Stochastic Dynamics: Mathematical Foundations of Diffusion and Score-Based Models

Rehan Ahmad

Department of Physics, LUMS

This independent study delves into the mathematical foundations of Score-Based and Diffusion Models through the lens of Stochastic Differential Equations (SDEs). By exploring key concepts such as the Fokker-Planck equation, Itô calculus, and reverse-time SDEs, it aims to build a comprehensive understanding of how these frameworks enable generative modeling.

Contents

1 Introduction

1.1 Background and Relevance

The mathematical study of Stochastic Differential Equations (SDEs) underpins the development of Score-Based Models and Diffusion Models which have gained prominence in generative modeling due to their ability to approximate complex distributions and perform sampling through reverse-time processes. By leveraging the Fokker-Planck equation, Itô calculus, and concepts like Variance Preserving and Variance Exploding SDEs, these models redefine generative tasks by connecting probabilistic distributions with stochastic processes.Understanding these foundations not only elucidates how data perturbations and score estimations work but also highlights the deep connections between physics-inspired models and modern machine learning.

2 Introduction to Stochastic Differential Equations

2.1 From ODEs to SDEs

Imagine you are modeling the growth of a population $x(t)$ of bacteria. In a deterministic setting, we might use the following differential equation to model the growth of the population:

$$\frac{dx}{dt} = rx \tag{2.1}$$

where r is the growth rate. This equation tells us that the rate of change of the population is proportional to its current size. Given an initial population $x(0)$, the future population at any time t is completely determined. The solution to this equation is trivial:

$$x(t) = x(0)e^{rt}. \tag{2.2}$$

Now, suppose the growth rate r is not constant but fluctuates due to random environmental factors such as changes in nutrient availability or temperature. In this case, the change in x is no longer deterministic. This is the subject of stochastic differential equations (SDEs).

To get a flavor of SDEs, let us begin with a linear ordinary differential equation with a deterministic driving force that describes a damped harmonic oscillator under an external force $f(t)$:

$$\frac{dx}{dt} = -\gamma x + f(t). \tag{2.3}$$

We can discretize this ODE and write it in terms of differentials:

$$\Delta x(t_n) = x(t_n)\Delta t + f(t_n)\Delta t \tag{2.4}$$

where $f(t_n)\Delta t$ is the driving term. The value of $x(t_n + \Delta t)$ is then given by:

$$x(t_n + \Delta t) = x(t_n) + \Delta x(t_n)$$

Substituting $\Delta x(t_n)$ from (2.4), we have:

$$x(t_n + \Delta t) = x(t_n) + x(t_n)\Delta t + f(t_n)\Delta t. \tag{2.5}$$

If we know the value of x at $t = 0$, then:

$$x(\Delta t) = x(0)(1 + \Delta t) + f(0)\Delta t. \tag{2.6}$$

What we are particularly interested in is the scenario where the driving term $f(t_n)\Delta t$ becomes random at each time step t_n. This involves replacing $f(t_n)$ with a random variable y_n at each t_n. As a result, the difference equation (2.4) transforms

into the following:

$$\Delta x(t_n) = x(t_n)\Delta t + y_n \Delta t \tag{2.7}$$

where $y_n \sim P(Y = y_1, \ldots, y_i)$.

This is called a *stochastic difference equation*. It states that at each time t_n, we pick a value for the random variable y_n sampled from its probability density $P(Y)$ and add $y_n \Delta t$ to $x(t_n)$. Consequently, we can no longer predict the exact value of x at some future time T in advance. Instead, $x(T)$ depends on the cumulative effect of all random increments y_n up to time T, which are determined sequentially as the process evolves.

The solution for x at time Δt is:

$$x(\Delta t) = x(0)(1 + \Delta t) + y_0 \Delta t \tag{2.8}$$

Thus, $x(\Delta t)$ is now a random variable. If the initial condition $x(0)$ is fixed (i.e., not random), then $x(t)$ can be expressed as a linear transformation of the random variable y_0, which represents the noise increment over the interval $[0, \Delta t]$. On the other hand, if $x(0)$ is also a random variable, then $x(t)$ is a linear combination of two random variables: the initial condition $x(0)$ and the noise increment y_0.

When we proceed to the next time step to calculate $x(2\Delta t)$, this value depends on $x(0)$, y_0, and the noise increment for the second step y_1. At each time step, the solution of the stochastic difference equation $x(t_n)$ is a random variable, and this random variable evolves as time progresses. Thus, solving a stochastic difference equation requires determining the probability density of $x(t_n)$ for all future times t_n. This process involves deriving the probability density of $x(t_n)$ from:

- The probability densities of the noise increments y_n.
- The probability density of $x(0)$, if $x(0)$ is random.

Stochastic differential equations (SDEs) are obtained by taking the limit $\Delta t \to 0$ in stochastic difference equations. In this continuous-time framework, the solution to an SDE is characterized not by a single trajectory but by a probability density function that describes the value of x at future times t. Just as with ordinary (deterministic) differential equations, finding a closed-form solution to an SDE is not always possible. However, in many simple cases, explicit solutions can be derived. For more complex systems, numerical methods or approximations are typically employed to study the evolution of the probability density over time.

In addition to obtaining the probability density for x at future times t_n, we can also ask how x evolves with time given a specific set of values for the random increments y_n.

This focuses on two key aspects:

- The probability density of x at specific times.
- The sample paths of x, which describe specific realizations of its evolution under different noise realizations.

Realization of the Noise : Let y_n represent the random noise increments (e.g., increments of a Wiener process) over discrete time intervals $[\Delta t, 2\Delta t, \dots]$. A realization of the noise is a specific set of sampled values $\{y_1, y_2, \dots, y_n\}$ drawn from the noise's probability density $y_n \sim N(0, \Delta t)$

Sample Path of $x(t)$: A sample path is a trajectory of $x(t)$ over time that corresponds to a specific realization of the noise. The full solution to the SDE is the collection of all possible sample paths, along with their associated probabilities. Mathematically, this corresponds to a probability measure over the space of functions $x(t)$. This full solution is rarely needed in practice. Instead, we focus on:

- The marginal probability density $P(x, t)$: The probability density of $x(t)$ at each time t, obtained from the Fokker-Planck equation associated with the SDE.
- Correlation Functions: Quantities like $\mathbb{E}[x(t)x(t')]$, which describe how $x(t)$ at one time is statistically related to $x(t')$ at another.

2.2 Introduction to Weiner Increments

By "Gaussian noise," we mean that each of the random increments $y_0\Delta t$ has a Gaussian probability density. Specifically, $y_n \sim \mathcal{N}(0, \sigma^2)$. First, consider the simplest stochastic difference equation where the increment of x consists solely of the random increment $y_n\Delta t$. This reduces equation (2.7) to:

$$\Delta x(t_n) = y_n \Delta t \tag{2.9}$$

In literature, Gaussian noise is often referred to as *Wiener noise*, and the random increment is expressed as:

$$\Delta W_n = y_n \Delta t \tag{2.10}$$

The discrete differential equation for x can thus be written as:

$$\Delta x(t_n) = \Delta W_n \tag{2.11}$$

Each Wiener increment ΔW_n is independent of the others and has the same probability density:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\Delta W)^2}{2\sigma^2}} \tag{2.12}$$

This density is Gaussian with zero mean, and we set the variance as $\sigma^2 = V = \Delta t$. This choice is critical and we will explain the reason for it in the next section. For simplicity, we often denote a Wiener increment in a given time step Δt as ΔW, without referencing the subscript n, since all increments share the same distribution and are independent.

Solving the Difference Equation for x:

We can solve the difference equation (2.11) for x by starting with $x(0) = 0$ and repeatedly adding Δx. The solution is:

$$x_n \equiv x(n\Delta t) = \sum_{i=1}^{n-1} \Delta W_i \tag{2.13}$$

The probability density of x_n can now be calculated. Since the sum of Gaussian random variables is also Gaussian, the probability density of x_n is Gaussian (a property proved in most introductory books on Statistics). The mean and variance of x_n are the sums of the means and variances of ΔW_i, respectively, because the ΔW_i are independent. Thus:

$$\mu = \langle x_n \rangle = 0 \tag{2.14}$$

$$V(x_n) = n\sigma^2 = n\Delta t \tag{2.15}$$

The solution to the difference equation is then:

$$P(x_n) = \frac{1}{\sqrt{2\pi V}} e^{-\frac{x_n^2}{2V}} = \frac{1}{\sqrt{2\pi n\Delta t}} e^{-\frac{x_n^2}{2n\Delta t}} \tag{2.16}$$

From Difference Equation to Differential Equation:

Transforming the difference equation (2.11) into a differential equation gives the simplest SDE:

$$dx = dW \tag{2.17}$$

To solve this, consider T future steps with N discrete time steps and take the limit as $N \to \infty$:

$$x(T) = \lim_{N\to\infty} \sum_{i=0}^{N-1} \Delta W_i = \int_0^T dW(t) = W(T) \tag{2.18}$$

Here, the stochastic integral $W(T) = \int_0^T dW(t)$ is defined as the *limit of the sum of all increments of the Wiener process.* For $x(T)$, the probability density is Gaussian because it sums independent Gaussian variables. The mean is zero since each variable has zero mean. The variance of $x(T)$ does not require the limit $N \to \infty$ because N factors out:

$$V(x(T)) = \sum_{i=1}^{N-1} V[\Delta W_i] = \sum_{i=1}^{N-1} \Delta t = N\Delta t = N(T/N) = T \tag{2.19}$$

Thus, the probability density of $W(T)$ is:

$$P(W(T)) = P(x(T)) = P(x,T) = \frac{1}{\sqrt{2\pi T}} e^{-\frac{x^2}{2T}} \tag{2.20}$$

Definitions: The Wiener process is defined as $W(T)$, while dW represents an increment of the Wiener process:

$$W(T) = \int_0^T dW \tag{2.21}$$

Note: While we often refer to dW as the Wiener process, the Wiener process is technically $W(T)$, and dW is its increment.

2.3 The variance of a wiener increment must satisfy Δt

So far, when considering Wiener increments, we have assumed that the variance of Wiener increments is Δt, which led to the fact that the variance of $x(T)$ is proportional to T. It turns out that setting the variance to any other value leads to an unphysical system. To show this, we set $V[\Delta W(\Delta t)] = \Delta t^\alpha$ and calculate the variance of $x(T)$ once again:

$$V(x(T)) = \sum_{i=1}^{N-1} V[\Delta W_i] = N(\Delta t)^\alpha = N\left(\frac{T}{N}\right)^\alpha = N^{(1-\alpha)}T^\alpha \tag{2.22}$$

Now we take the continuum limit $N \to \infty$ to obtain a stochastic differential equation. When $\alpha > 1$, we have:

$$\lim_{N\to\infty} V(x(T)) = T^\alpha \lim_{N\to\infty} N^{1-\alpha} = 0 \tag{2.23}$$

And when $\alpha < 1$, we have:

$$\lim_{N\to\infty} V(x(T)) = T^\alpha \lim_{N\to\infty} N^{1-\alpha} = \infty \tag{2.24}$$

Neither of these results make sense for obtaining a stochastic differential equation that describes real systems driven by noise. Thus, we are forced to choose $\alpha = 1$ and hence $V[\Delta W(\Delta t)] \propto \Delta t$.

2.4 General SDE Form and an Example of SDE

Given that we have introduced the fundamentals of SDEs, let us now introduce a bit of formalism. A general SDE for a single variable X_t can be written as:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \tag{2.25}$$

Where:

- X_t is the stochastic process.
- $\mu(X_t, t)$ is the drift term, representing the deterministic part of the change.

- $\sigma(X_t, t)$ is the diffusion term, representing the stochastic part of the change.
- dW_t is the increment of the Wiener process.

Since the variance of dW_t must be proportional to dt, and any constant of proportionality can always be absorbed into $\sigma(X_t, t)$, the variance of dW_t is defined to be equal to dt. Therefore, the probability density of dW_t is:

$$P(dW) = \frac{1}{\sqrt{2\pi dt}} e^{-\frac{(dW)^2}{2dt}} \tag{2.26}$$

Building on the above, let us consider a simple SDE of the form:

$$dX = e^t dW \tag{2.27}$$

Where $W(t)$ is a Wiener process. This SDE can be expressed as:

$$dX = f(t)dW \tag{2.28}$$

To solve this equation, we integrate both sides from 0 to t:

$$X(t) - X(0) = \int_0^t e^s dW(s) \tag{2.29}$$

Assuming the initial condition $X(0) = 0$, we have:

$$X(t) = \int_0^t e^s dW(s) \tag{2.30}$$

This integral represents the accumulation of the stochastic process e^s weighted by the increments of the Wiener process over $[0, t]$.

Mean of $X(t)$: To understand the properties of $X(t)$, let us first calculate its mean. Since the integrand e^s and the Wiener process increments $dW(s)$ are independent and the expectation of $dW(s)$ is zero:

$$\mathbb{E}[X(t)] = \mathbb{E}\left[\int_0^t e^s dW(s)\right] = 0 \tag{2.31}$$

Variance of $X(t)$: The variance of $X(t)$ can be calculated using Itô isometry (a result we will prove later, :

$$V(X(t)) = \mathbb{E}\left[\left(\int_0^t e^s dW(s)\right)^2\right] = \int_0^t (e^s)^2 ds = \int_0^t e^{2s} ds \tag{2.32}$$

Evaluating the integral, we obtain:

$$V(X(t)) = \frac{e^{2t} - 1}{2} \tag{2.33}$$

3 Ito Calculus

3.1 Ito's Rule

Consider the solution to the simple differential equation:

$$dx = -\theta x \, dt \tag{3.1}$$

In elementary courses on Ordinary Differential Equations (ODEs), the solution to this equation is straightforward. The deterministic nature of this equation ensures a well-defined trajectory for $x(t)$. However, when we introduce stochastic elements, the rules for solving such equations become less straightforward, as we will see.

To understand this, let us first revisit the solution to the deterministic equation (3.1) . The value of x at time $t + dt$ is the value at time t plus the infinitesimal change dx:

$$x(t + dt) = x(t) - \theta x(t) dt = (1 - \theta dt) x(t) \tag{3.2}$$

From the Taylor series, $e^{\theta dt} \approx 1 + \theta dt$. We can therefore write the equation for $x(t + dt)$ as:

$$x(t + dt) = e^{-\theta dt} x(t) \tag{3.3}$$

This tells us that to move x from time t to $t + dt$, we merely have to multiply $x(t)$ by the factor $e^{-\theta dt}$. So to move by two lots of dt, we simply multiply by this factor twice:

$$x(t + 2dt) = e^{-\theta dt} x(t + dt) = e^{-\theta dt} [e^{-\theta dt} x(t)] = e^{-\theta \cdot 2dt} x(t). \tag{3.4}$$

To obtain $x(t + \tau)$, all we have to do is apply this relation repeatedly. Let us say that $dt = \tau/N$ for N as large as we want. Thus, dt is a small but finite time-step, and we can make it as small as we want. That means that to evolve x from time t to $t + \tau$, we can apply (3.3) N times:

$$x(t + \tau) = \left(e^{-\theta dt}\right)^N x(t) = e^{-\theta \sum_{n=1}^{N} dt} x(t) = e^{-\theta N dt} x(t) = e^{-\theta \tau} x(t) \tag{3.5}$$

is the solution to the differential equation. If γ is a function of time, so that the equation becomes:

$$dx = -\theta(t) x \, dt \tag{3.6}$$

As before, we set $dt = \tau/N$ so that it is a small finite-time step, but this time we have to explicitly take the limit as $N \to \infty$ to obtain the solution to the differential

equation:

$$\begin{aligned} x(t+\tau) &= \lim_{N\to\infty} \prod_{n=1}^{N} \left(e^{-\theta(t+ndt)dt}\right) x(t) \\ &= \lim_{N\to\infty} e^{-\sum_{n=1}^{N} \theta(t+ndt)dt} x(t) \\ &= e^{-\int_t^{t+\tau} \theta(t)dt} x(t) \end{aligned} \tag{3.7}$$

Notice that we were able to solve the equation (3.3) using $e^{\alpha dt} \approx 1 + \alpha dt$. The approximation $e^{\alpha dt} \approx 1 + \alpha dt$ works because the terms in the power series expansion for $e^{\alpha dt}$ that are second-order or higher in dt ($dt^2, dt^3, \ldots$) will vanish in comparison to dt as $dt \to 0$. The result of being able to ignore terms that are second-order and higher in the infinitesimal increment leads to the usual rules for differential equations. (It also means that any equation we write in terms of differentials dx and dt can alternatively be written in terms of derivatives.)

However, we will now spend some time showing that as $dt \to 0$, $(dW)^2$ does not tend to 0. We must therefore learn a new rule for the manipulation of stochastic differential equations.

To examine whether $(dW)^2$ makes a non-zero contribution to the solution, we sum $(dW)^2$ over all the time-steps for a finite time T. To do this, we will return to a discrete description so that we can explicitly write down the sum and then take the continuum limit. That is, we examine the limit:

$$\Delta X = \lim_{N\to\infty} \sum_{i=1}^{N} (dW_i)^2 \tag{3.8}$$

The first thing to note is that the expectation value of $(\Delta W)^2$ is equal to the variance of ΔW, because $\langle \Delta W \rangle = 0$. From $\mathrm{Var}(\Delta W) = \langle (\Delta W)^2 \rangle - \langle \Delta W \rangle^2$, we have:

$$\langle \Delta W^2 \rangle = \Delta t \tag{3.9}$$

This tells us immediately that the expectation value of $(\Delta W)^2$ does not vanish with respect to the time-step Δt, and so the sum of these increments will not vanish when we sum over all the time-steps and take the infinitesimal limit (because we would have a Gaussian on our hands with a variance Δt). In fact, the expectation value of the sum of all the increments $(dW)^2$ from 0 to T is simply T:

$$\langle \int_0^T (dW)^2 \rangle = \int_0^T \langle (dW)^2 \rangle = \int_0^T dt = T \tag{3.10}$$

Claim: The Integral of $(\Delta W)^2$ is Deterministic:

In order to show that the integral of $(\Delta W)^2$ is deterministic, we will look at the variance $\int_0^T (dW)^2$ and show that it is zero. To do this, we will first argue about the

differential $(\Delta W)^2$ itself. We found that the expectation of $(\Delta W)^2$ is proportional to Δt, and so we have a strong hunch that the variance of $(\Delta W)^2$ must be proportional to $(\Delta t)^2$. *Hint*: $\text{Var}(x) = \mathbb{E}[x^2] - (\mathbb{E}[x])^2$. Let us now explicitly calculate the variance of $(\Delta W)^2$ using the probability density for ΔW:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-\frac{(\Delta W)^2}{2\Delta t}} \tag{3.11}$$

Using the change of variable theorem, we now find the probability distribution for $(\Delta W)^2$. We define $z = (\Delta W)^2$. For this, we have two cases: $\Delta W = \sqrt{z}$ and $\Delta W = -\sqrt{z}$. The absolute value of the derivative $\frac{d\Delta W}{dz}$ for $z = (\Delta W)^2$ is:

$$\left|\frac{d\Delta W}{dz}\right| = \left|\frac{d}{dz}(\pm\sqrt{z})\right| = \frac{1}{2\sqrt{z}}.$$

Using the change of variables formula for PDFs, we have:

$$f_z(z) = f_{\Delta W}(\sqrt{z}) \left|\frac{d\Delta W}{dz}\right| + f_{\Delta W}(-\sqrt{z}) \left|\frac{d\Delta W}{dz}\right|.$$

Since $f_{\Delta W}(x)$ is even, $f_{\Delta W}(-\Delta W) = f_{\Delta W}(\Delta W)$. Thus:

$$f_z(z) = \frac{f_{\Delta W}(\sqrt{z})}{\sqrt{z}}.$$

Substituting the original PDF, we obtain:

$$f_z((\Delta W)^2) = \frac{1}{\sqrt{2\pi\Delta t z}} e^{-\frac{z}{2\Delta t}}. \tag{3.12}$$

This is a **chi-squared distribution** with one degree of freedom, scaled by a factor of $2\Delta t$. The variance of the distribution is:

$$V((\Delta W)^2) = 2(\Delta t)^2 \tag{3.13}$$

The variance of the sum of all the $(\Delta W)^2$ is:

$$V\left[\sum_{n=1}^{N-1}(\Delta W)^2\right] = \sum_{n=1}^{N-1} V((\Delta W)^2) = \sum_{n=1}^{N-1} 2(\Delta t)^2 = 2N\left(\frac{T}{N}\right)^2 = \frac{2T^2}{N} \tag{3.14}$$

Finally, we can perform what we set out to do, namely evaluating the integral in equation (3.8):

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N-1} V((\Delta W)^2)\right] = \lim_{N\to\infty}\frac{2T^2}{N} = 0. \tag{3.15}$$

Since the integral of all the $(dW)^2$ is deterministic, it is equal to its mean T. That is:

$$\int_0^T (dW)^2 = T = \int_0^T dt \tag{3.16}$$

Thus, we have the surprising result:

$$dW^2 = dt \tag{3.17}$$

This is officially known as ***Ito's rule***. It is the fundamental rule for solving stochastic differential equations that contain Gaussian noise.

3.2 Ito's Lemma

Due to the Ito rule established in equation (3.17), we will have to keep all terms that are first order in dt and dW, as well as all terms that are second order in dW. In fact, wherever we find terms that are second order in dW, we can simply replace them with dt. To see how this works, consider a simple example in which we want to know the differential equation for:

$$y = x^2 \tag{3.18}$$

In that case:

$$\begin{aligned} dy &= y(t+dt) - y(t) \\ &= x(t+dt)^2 - x(t)^2 \\ &= (x+dx)^2 - x^2 \\ &= x^2 + 2xdx + (dx)^2 - x^2 \\ &= 2xdx + (dx)^2 \end{aligned} \tag{3.19}$$

Had x been deterministic, $(dx)^2$ would vanish in the continuum limit, and we would have the usual rule of calculus:

$$dy = 2xdx \quad \text{or} \quad \frac{dy}{dx} = 2x \tag{3.20}$$

However, if X is a random variable obeying the stochastic differential equation:

$$dX = fdt + gdW \tag{3.21}$$

then the equation (3.17) becomes the following:

$$\begin{aligned} dY &= 2xdX + (dX)^2 \\ &= 2x(fdt + gdW) + (fdt + gdW)^2 \end{aligned} \tag{3.22}$$

Expanding the square term, we have:

$$(f dt + g dW)^2 = f^2 dt^2 + g^2 dW^2 + 2fg dt dW.$$

Just like in normal calculus, we can ignore cross-differentials like $dt\, dW$ and dt^2.

$$(f dt + g dW)^2 \approx g^2 dW^2.$$

We then have:

$$dY = 2fX dt + 2xg dW + g^2 dW^2. \tag{3.23}$$

Using Ito's rule in equation (3.17):

$$dY = (2fX + g^2)dt + 2xg dW \tag{3.24}$$

This is Ito's rule in action.

Ito's Lemma: Fortunately, there is a simple way to calculate the increment of any nonlinear function $y(X)$ in terms of the first and second powers of the increment of X, where X is a random variable. All we have to do is use the Taylor series expansion for $y(X)$, truncated at the second term:

$$dy(X) = \frac{dy}{dX} dX + \frac{1}{2} \frac{d^2 y}{dX^2} (dX)^2 \tag{3.25}$$

If y is also an explicit function of time as well as X, then this becomes:

$$dy(t, X) = \frac{\partial y}{\partial X} dX + \frac{\partial y}{\partial t} dt + \frac{1}{2} \frac{\partial^2 y}{\partial X^2} (dX)^2 \tag{3.26}$$

3.3 Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck (OU) process is a type of stochastic process used to model mean-reverting behavior. It was introduced by Leonard Ornstein and George Eugene Uhlenbeck in 1930 to describe the velocity of a particle undergoing Brownian motion under the influence of friction.

Understanding Mean-Reverting Behavior

Imagine a particle diffusing in a liquid where there is some friction or restoring force that pulls the particle back towards a central point (mean position) rather than allowing it to drift indefinitely. In this context, the position X_t of the particle at time t can be described by a stochastic differential equation governed by two terms, $\theta(\mu - X_t)$ and σdW_t:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t \tag{3.27}$$

Here:

- θ is the strength of the restoring force, causing the particle to revert to the mean position μ,
- σ is the volatility and represents the intensity of the random fluctuations.

To understand how mean-reverting dynamics would play out, suppose the particle starts at the position X_t far from the mean μ. For example, if μ is at the origin (0) and the particle is initially at $X_0 = 5$, the term $\theta(\mu - X_t)dt$ acts as a restoring force that pulls the particle back towards the mean position μ. The larger θ is, the stronger the pull back to μ. Despite the mean-reverting force, the particle is subject to random fluctuations due to the noise term σdW_t. This represents the random kicks from the surrounding fluid molecules.

Visually, processes obeying (3.27) look like the following:

Include plot illustrating sample paths of the Ornstein-Uhlenbeck process

Here we can see that even though two of the three sample paths start far away from the mean, they quickly converge to a region around the mean. Once in the vicinity of the mean, they move about it in arcs through the momentum factor.

Solving the Ornstein-Uhlenbeck Equation

First, we define a new random variable Y_t:

$$Y_t = X_t - \mu \tag{3.28}$$

The differential element of dY_t is given by:

$$\begin{aligned} dY_t &= dX_t \\ &= \theta(\mu - X_t)dt + \sigma dW_t \\ &= -\theta(X_t - \mu)dt + \sigma dW_t \\ &= -\theta Y_t dt + \sigma dW_t. \end{aligned} \tag{3.29}$$

Equation 3.29 is the reason why we say that Orstein-Uhlenbeck equation is governed by "additive noise". The term "additive noise" refers to the fact that **the noise does not itself depend on** Y_t, but is merely added to any other terms that appear in the equation for dx. This equation is called the Ornstein–Uhlenbeck equation, and its solution is called the Ornstein–Uhlenbeck process

The next step is to recognize that we are equating the derivative of a random variable with itself:

$$\begin{aligned} dY_t &\propto \theta Y_t dt \\ Y_t &\propto Y_t e^{\theta t} \end{aligned}$$

To look for simplifications, we thus define another random variable Z_t as a function of Y_t :

$$Z_t = f\left(t, \theta, Y_t\right) = e^{\theta t} Y_t \tag{3.30}$$

Using Ito's Lemma in equation (3.26), it follows:

$$df\left(t, Y_t\right) = \left(\frac{\partial f}{\partial Y_T}\right) dY_t + \left(\frac{\partial f}{\partial t}\right) dt + \frac{1}{2}\left(\frac{d^2 f}{dY_T^2}\right) \left(dY_T\right)^2 \tag{3.31}$$

Where note that $\frac{d^2 f}{dY_T^2} = \frac{d^2}{dY_T^2}\left[e^{\theta t} Y_t\right] = 0$. Thus, Ito formula assumes the following form in our case:

$$df\left(\theta, Y_t\right) = \left(\frac{\partial f}{\partial Y_T}\right) dY_t + \left(\frac{\partial f}{\partial t}\right) dt \tag{3.32}$$

Lets apply this,

$$dZ\left(\theta, Y_t\right) = \theta e^{\theta t} Y_t dt + e^{\theta t} dY_t$$

Plugging equation 3.29,

$$\begin{aligned} dZ(\theta, Y_t) &= \theta e^{\theta t} Y_t dt + e^{\theta t}\left(-\theta Y_t dt + \sigma dW_t\right) \\ &= e^{\theta t} \sigma dW_t \end{aligned} \tag{3.33}$$

Equation (3.33) is easy to solve. To do so we merely sum all the stochastic increments dW over a finite time t , noting that each one is multiplied by $e^{\theta t}\sigma$. Thus, the integral form is.

$$\begin{aligned} Z_t &= Z_s + \int_s^T dZ_t \\ &= Z_s + \sigma \int_s^T e^{\theta t} dW_t \end{aligned} \tag{3.34}$$

Where S is the start of the integration through time. Now that we found a solution to the random variable Z_t it is time to go back through the substitutions to find the solution to X_t. In order to achieve that we first reverse the exponential component in the relationship between Y_t and Z_t :

$$
\begin{aligned}
Y_t &= e^{-\theta t} Z_t \\
Y_T &= e^{-\theta T} \left(Z_S + \sigma \int_S^T e^{\theta t} dW_t \right) \\
Y_T &= e^{-\theta T} \left(e^{kS} Y_S + \sigma \int_S^T e^{\theta t} dW_t \right) \\
Y_T &= e^{-\theta (T-S)} Y_S + \sigma \int_S^T e^{\theta (t-T)} dW_t. \qquad (3.35)
\end{aligned}
$$

Finally plugging $Y_t = X_t - \mu$:

$$
\begin{aligned}
X_T - \mu &= e^{-\theta (T-S)} (X_S - \mu) + \sigma \int_S^T e^{\theta (t-T)} dW_t \\
X_T &= \mu + e^{-\theta (T-S)} (X_S - \mu) + \sigma \int_S^T e^{\theta (t-T)} dW_t
\end{aligned}
$$

Starting from $S = 0$, we obtain:

$$
X_T = \mu + e^{-\theta T} (X_0 - \mu) + \sigma \int_{s=0}^T e^{-\theta (T-s)} dW_s \qquad (3.36)
$$

To completely determine X_T in 3.36, note that the stochastic integral represents the sum of Gaussian random variables. Thus all we need to do is calculate its mean and variance:

$$
\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}\left[\mu + e^{-\theta T} (X_0 - \mu) + \sigma \int_{s=0}^T e^{-\theta (T-t)} dW_s \right] \\
\mathbb{E}[X_t] &= \mu + e^{-\theta T} (X_0 - \mu) + \mathbb{E}\left[\sigma \int_{s=0}^T e^{-\theta (T-t)} dW_s \right]
\end{aligned}
$$

Where $\int_{S=0}^T e^{-\theta (T-t)} dW_s$ is a Weiner process and thus, $\mathbb{E}\left[\sigma \int_{S=0}^T e^{-\theta (T-t)} dW_S \right] = 0$.

$$
\mathbb{E}[X_t] = \mu + e^{-\theta T} (X_0 - \mu) \qquad (3.37)
$$

Let us now compute the variance.

$$
\begin{aligned}
\mathbb{V}(Z_t) &= \mathbb{E}\left[(X_t - \mathbb{E}[X_t])^2 \right] \\
\mathbb{V}(Z_t) &= \mathbb{E}\left[\left(\mu + e^{-\theta T} (X_0 - \mu) + \sigma \int_{s=0}^T e^{-\theta (T-s)} dW_s - \mathbb{E}[X_t] \right)^2 \right]
\end{aligned}
$$

Where plugging in (3.37) inside the expression,

$$
\mathbb{V}(X_t) = \mathbb{E}\left[\left(\sigma \int_{s=0}^T e^{-\theta (T-s)} dW_s \right)^2 \right]
$$

From Ito isometry, $dW^2 = dt$,

$$\begin{aligned}
\mathbb{V}(X_t) &= \sigma^2 \int_{s=0}^{T} e^{-2\theta(T-s)} ds \\
\mathbb{V}(X_t) &= \sigma^2 \left[\frac{1}{2\theta} e^{-2\theta(T-s)} \right]_{s=0}^{t} \\
\mathbb{V}(X_t) &= \frac{\sigma^2}{2\theta} \left(1 - e^{-2\theta t}\right). \qquad (3.38)
\end{aligned}$$

Applying the limit $t \to \infty$ allows us to recover the stationary distribution:

$$\lim_{t\to 0} \mathbb{E}[X_t] = \lim_{t\to\infty} \mu + e^{-\theta t}(X_0 - \mu) = \mu \qquad (3.39)$$

$$\lim_{t\to 0} \mathbb{V}[X_t] = \lim_{t\to\infty} \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta t}\right) = \frac{\sigma^2}{2\theta} \qquad (3.40)$$

3.4 The Full Linear Stochastic Equation

The general stochastic linear equation reads as following:

$$dS_t = -\mu_t S_t \, dt + \sigma_t S_t \, dW_t \qquad (3.41)$$

where $S_t(W_t)$ do not have explicit time dependence. This equation is called the **Geometric Brownian Motion**. We can write the above as following:

$$\frac{dS_t}{S_t} = \mu_t \, dt + \sigma_t \, dW_t \qquad (3.42)$$

Notice that as S_t approaches zero, so does the change (else the LHS blows up). This effectively limits S_t to positive values, $S_t > 0$.

INSERT FIGURE HERE

The question is as so often with differential equations, whether there exists an analytic solution. In order to show this analytic solution we will examine the quantity $\frac{dS_t}{S_t}$ and apply the stochastic version of the log-derivative trick. The quantity $\frac{dS_t}{S_t}$ has striking similarity to:

$$\frac{\partial \ln S(x)}{\partial x} = \frac{1}{S(x)} \frac{\partial S(x)}{\partial x} \qquad (3.43)$$

But since we are working with stochastic processes, we can't apply regular calculus to derive such a stochastic process but use Ito's lemma (3.26):

$$d \ln S_t = \frac{\partial \ln S_t}{\partial t} dt + \frac{\partial \ln S_t}{\partial S_t} dS_t + \frac{1}{2} \frac{\partial^2 \ln S_t}{\partial S_t^2} dS_t^2 \qquad (3.44)$$

For $\ln S_t$, we calculate the derivatives:

$$\frac{\partial S_t(W_t)}{\partial t} = 0, \quad \frac{\partial \ln S_t}{\partial S_t} = \frac{1}{S_t}, \quad \frac{\partial^2 \ln S_t}{\partial S_t^2} = -\frac{1}{S_t^2}.$$

The above can then be written as:

$$d \ln S_t = \frac{1}{S_t} dS_t - \frac{1}{2} \frac{1}{S_t^2} dS_t^2. \tag{3.45}$$

Plugging in equation (3.41),

$$d \ln S_t = \frac{1}{S_t} dS_t - \frac{1}{2} \frac{1}{S_t^2} \left(-\mu_t S_t \, dt + \sigma_t S_t \, dW_t\right)^2. \tag{3.46}$$

Expanding the term:

$$\left(-\mu_t S_t \, dt + \sigma_t S_t \, dW_t\right)^2 = \mu_t^2 S_t^2 \, dt^2 + \sigma_t^2 S_t^2 \, dW_t^2 - 2\mu_t S_t \sigma_t S_t \, dW_t \, dt.$$

Ignoring the terms dt^2 and $dW_t \, dt$:

$$\left(-\mu_t S_t \, dt + \sigma_t S_t \, dW_t\right)^2 \approx \sigma_t^2 S_t^2 \, dW^2.$$

Using Ito's rule and (3.45):

$$\begin{aligned} d \ln S_t &= \frac{1}{S_t} dS_t - \frac{1}{2} \sigma_t^2 \, dt \\ \frac{1}{S_t} dS_t &= d \ln S_t + \frac{1}{2} \sigma_t^2 \, dt. \end{aligned} \tag{3.47}$$

Substituting this into equation (3.42):

$$\begin{aligned} \mu_t dt + \sigma_t dW_t &= d \ln S_t + \frac{1}{2} \sigma_t^2 dt \\ \int_0^t d \ln S_t &= \int_0^t \mu_s \, ds + \int_0^t \frac{1}{2} \sigma_s^2 \, ds + \int_0^t \sigma_s \, dW_s \\ \ln S_t - \ln S_0 &= \mu_t t - \frac{1}{2} \sigma_t^2 t + \sigma_t W_t \\ \ln \frac{S_t}{S_0} &= \mu_t t - \frac{1}{2} \sigma_t^2 t + \sigma_t W_t \\ S_t &= S_0 e^{\mu_t t - \frac{1}{2} \sigma_t^2 t + \sigma_t W_t}. \end{aligned} \tag{3.48}$$

3.5 The stationary auto-correlation function $g(\tau)$

For a stochastic process, it is often useful to know how correlated the values of the process are at two different times. This will tell us how long it takes the process to forget the value it had at some earlier time. We are therefore interested in calculating the correlation coefficient:

$$C_{X(t)X(t+\tau)} = \frac{\langle X(t)X(t+\tau)\rangle - \langle X(t)\rangle\langle X(t+\tau)\rangle}{\sqrt{\mathbb{V}(X(t))\mathbb{V}(X(t+\tau))}} \tag{3.49}$$

for an arbitrary time difference τ. As an illustration we calculate this for the Wiener process. We know already that $\mathbb{V}(W(t)) = t$ and thus $\mathbb{V}(W(t+\tau)) = t+\tau(\mathrm{W(t+\tau)}) = \mathrm{t}+\tau$. We can calculate the correlation $\langle W(t)W(t+\tau)\rangle$ in the following way:

$$\begin{aligned}
\langle W(t)W(t+\tau)\rangle &= \left\langle \int_0^t dW \int_0^{t+\tau} dW \right\rangle = \left\langle \int_0^t dW \left(\int_0^t dW + \int_t^{t+\tau} dW \right) \right\rangle \\
\langle W(t)W(t+\tau)\rangle &= \left\langle \left(\int_0^t dW \right)^2 + \int_0^t dW \int_t^{t+\tau} dW \right\rangle \\
\langle W(t)W(t+\tau)\rangle &= \left\langle \left(\int_0^t dW \right)^2 \right\rangle + \left\langle \int_0^t dW \int_t^{t+\tau} dW \right\rangle \\
\langle W(t)W(t+\tau)\rangle &= \left\langle (W(t))^2 \right\rangle + \left\langle \int_0^t dW \right\rangle \left\langle \int_t^{t+\tau} dW \right\rangle \\
\langle W(t)W(t+\tau)\rangle &= \left\langle (W(t))^2 \right\rangle + \langle W(t)\rangle\langle W(\tau)\rangle \\
\langle W(t)W(t+\tau)\rangle &= t + 0 = t.
\end{aligned} \tag{3.50}$$

Thus, the correlation coefficient is given as following:

$$C_{X(t)X(t+\tau)} = \frac{t}{\sqrt{t(t+\tau)}} = \sqrt{\frac{1}{(1+\tau/t)}} \tag{3.51}$$

Where we used the fact that the random variables $\mathrm{A} = \int_0^t dW$ and $\mathrm{B} = \int_t^{t+\tau} dW$ are independent, which implies their correlation $\langle AB\rangle$ is just the product of their means, $\langle A\rangle\langle B\rangle$. As expected, the Wiener process at time $\mathrm{t}+\tau$ is increasingly independent of its value at an earlier time t as τ increases.

The meaning of auto-correlation function: The function:

$$\mathrm{g\,(t, t')} = \langle \mathrm{X(t)X\,(t')}\rangle \tag{3.52}$$

is often called the two-time correlation function or the auto-correlation function ("auto" because it is the correlation of the process with itself at a later time).

If the mean of the process X(t) is constant with time, and the auto-correlation function, $\mathrm{g(t, t+\tau)} = \langle \mathrm{X(t)X(t+\tau)}\rangle$ is also independent of the time, t , so that it depends only on the timedifference, τ, then X(t) is referred to as being "wide-sense"

stationary. In this case, the autocorrelation function depends only on τ, and we write:

$$g(\tau) = \langle \mathrm{X(t)X\,(t')} \rangle \tag{3.53}$$

The auto-correlation function for a wide-sense stationary process is always symmetric, so that $\mathrm{g}(-\tau) = \mathrm{g}(\tau)$. This is easily shown by noting that:

$$g(\tau) = \langle \mathrm{X(t)X(t-\tau)} \rangle = \langle \mathrm{X(t)X(t+\tau)} \rangle = g(\tau) \tag{3.54}$$

Calculating auto-correlation using the conditional Probability Density:

One can always calculate the correlation $\langle \mathbf{X}\,(\mathrm{t'})\,\mathbf{X}(\mathrm{t}) \rangle$ at two times t′ and $\mathrm{t} = \mathrm{t'} + \tau$, for some arbitrary process $X(t)$, so long as one has the joint probability density that the value of the process is x at time t and x ' and time t '. Let us define the probability density as,

$$P\,(x,t;x',t') = P\,(x,t \mid x',t')\,P\,(x',t') \tag{3.55}$$

The conditional probability is the probability density for X at time t , given that X has the value x ' at time t '. In fact, we already know how to calculate this, since it is the same thing that we have been calculating all along in solving stochastic differential equations: the solution to an SDE for X is the probability density for X at time t , given that its initial value at $\mathrm{t} = 0$ is x_0. To obtain the conditional probability in Eq. (14), all we need to do is solve the SDE for x , but this time with the initial time being t' rather than 0.

As an example, let us do this for the simplest stochastic equation, $\mathrm{d}\mathbf{X} = \mathrm{d}\mathbf{W}$. Solving the SDE means summing all the increments dW from time t′ to t , with the initial condition $X\,(t') = x'$. The solution is,

$$X(t) = x' + \int_{t'}^{t} dW = x' + W\,(t-t') \tag{3.56}$$

And this has the probability density,

$$P\,(x,t \mid x',t') = P(x,t) = \frac{e^{-(x-x')^2/[2(t-t')]}}{\sqrt{2\pi\,(t-t')}} \tag{3.57}$$

To calculate the joint probability density we now need to specify the density for X at time t′. If X started with the value 0 at time 0 , then at time t' the density for $\mathrm{X\,(t')}$ is just the density for the Wiener process, thus,

$$P\left(x',t'\right)=\frac{e^{-\frac{x'^2}{[2t^2]}}}{\sqrt{2\pi}t'} \tag{3.58}$$

Using equation 3.57 and 3.58, the joint probability density is,

$$P\left(w,t\mid w',t'\right)=\frac{e^{-\frac{(x-x')^2}{[2(t-t')]-\frac{x'^2}{[2t']}}}}{\sqrt{2\pi\left(t-t'\right)t'}} \tag{3.59}$$

And the correlation function is therefore,

$$\left\langle X\left(t'\right)X(t)\right\rangle=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xx'P\left(x,t;x',t'\right)dxdx'=t' \tag{3.60}$$

We then obtain the correlation coefficient between X(t) and X (t′) by dividing this by the square root of the product of the variances as above

4 Properties of Stochastic Processes

4.1 Wiener-Khinchin theorem

Let us define what we mean by the Fourier transform of a stochastic process. A stochastic process, $\mathbf{x}(\mathrm{t})$, has many possible sample paths. Therefore, we can think of $\mathbf{x}(\mathrm{t})$ as being described by a probability density over the whole collection of possible sample paths. Each one of these sample paths is a function of time, say $\boldsymbol{x}_\alpha(\mathrm{t})$, where α labels the different possible paths. Thus $\mathbf{x}(\mathrm{t})$ is actually a random function, whose possible values are the functions $\boldsymbol{x}_\alpha(t)$. Just as $\mathbf{x}(t)$ is a random function, whose values are the sample paths, we define the Fourier transform, $\mathbf{x}(\omega)$, as a random function whose values are the Fourier transforms of each of the sample paths. Thus the possible values of $\mathbf{x}(\omega)$ are the functions:

$$\mathbf{X}_\alpha(\omega)=\int_{-\infty}^{\infty}\mathbf{x}_\alpha(\mathrm{t})e^{-i2\pi\omega t}dt \tag{4.1}$$

For a stochastic signal, $\mathbf{x}(\mathrm{t})$, the total average energy in the signal is the average value of the instantaneous power, $\mathbf{x}^2(\mathrm{t})$, integrated over all time:

$$\mathrm{E}[\mathbf{x}(\mathrm{t})]=\int_{-\infty}^{\infty}\left\langle\mathbf{x}(\mathrm{t})^2\right\rangle\mathrm{dt} \tag{4.2}$$

With the total energy being conserved in the ω-space also:

$$\mathrm{E}[\mathbf{X}(\omega)]=\int_{-\infty}^{\infty}\left\langle\left|\boldsymbol{X}(\omega)\right|^2\right\rangle\mathrm{d}\omega \tag{4.3}$$

With this definition of the Fourier transform of a stochastic process, we can now

derive the proof that the autocorrelation and the energy spectrum of the stochastic process are Fourier pairs just like we did for deterministic functions.

$$\begin{aligned}
\langle |F(v)|^2 \rangle &= \left\langle \int_{-\infty}^{\infty} f(t) e^{-i2\pi\omega t} dt \int_{-\infty}^{\infty} f(t) e^{i2\pi\omega t} dt \right\rangle \\
\langle |F(v)|^2 \rangle &= \left\langle \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (f(t) f(t-\tau) dt) e^{-i2\pi\omega t} d\tau \right\rangle \\
\langle |F(v)|^2 \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle f(t) f(t-\tau) dt \rangle e^{-i2\pi\omega t} d\tau \\
\langle |F(v)|^2 \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t,\tau) e^{-i2\pi\omega t} d\tau
\end{aligned} \tag{4.4}$$

Where,

$$g(t,\tau) = \langle f(t) f(t+\tau) \rangle \tag{4.5}$$

Let us now define the average power of a stochastic signal in the same way as for a deterministic signal, but this time we take the expectation value of the process, so as to average the power both over time and over all realizations (sample paths) of the process. Thus the average power of a stochastic process is:

$$P[x(t)] = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} \left\langle x^2(t) \right\rangle dt \tag{4.6}$$

With the average power defined, it is useful to be able to power spectral density, S(v). This is defined as the power per unit frequency of a sample path of the process, averaged over all the sample paths:

$$P[Y(t)] = \int_{v_1}^{v_2} S(v) dv \tag{4.7}$$

It turns out that the power spectral density of a wide-sense stationary stochastic process $x(t)$ is the Fourier transform of the two-time auto-correlation function. Thus,

$$S(v) = \int_{-\infty}^{\infty} g(\tau) e^{-2\pi v \tau} d\tau, g(\tau) = \langle x(t) x(t+\tau) \rangle \tag{4.8}$$

This result is called the Wiener-Khinchin theorem.

4.2 Power Spectrum of Stochastic Processes:

Different types of noise are generated by different stochastic processes. The power spectrum of a noise signal is referred to using colors.

One type is white noise, which is when each component of the noise signal have a probability distribution with zero mean and finite variance, and are statistically independent. This results in a noise signal with spectral density that is even throughout

all frequencies (flat power spectral density). Note that the name is drawn from the white light as it contains all colors.

Another type is red noise or Brownian noise, which refers to noise resulting from Brownian motion. The spectral density of this type is inversely proportional to the frequency squared. Meaning, its power drastically decreases as its frequency increases (has more energy at low frequencies). Note that it is called red noise as it is analogous to red light which has a low frequency.

ADD FIGURE HERE

White Noise:

Consider a function $f(t)$ whose integral from $-\infty$ to ∞ is finite. The more sharply peaked f (that is, the smaller the smallest time interval containing the majority of its energy), then the less sharply peaked is its Fourier transform. Similarly, the broader a function, then the narrower is its Fourier transform. Now consider a stochastic process $x(t)$. If the auto-correlation function $g(\tau) = \mathrm{x(t)x(t+\tau)}$ drops to zero very quickly as $|\tau|$ increases, then the power spectrum of the process x(t) must be broad, meaning that x(t) contains high frequencies. This is reasonable, since if a process has high frequencies it can vary on short time-scales, and therefore become uncorrelated with itself in a short time.

We know that the derivative of the Wiener process does not exist. However, there is a sense in which the auto-correlation of this derivative exists. This can be useful as a calculational tool. For the sake of argument let us call the derivative of the Wiener function $\xi(t)$. Since the increments of the Wiener process in two consecutive time intervals dt are independent of each other, ξ(t) must be uncorrelated with itself whenever the time separation is greater than zero. Thus we must have $\langle \xi(t)\xi(t+\tau)\rangle = 0$ if $\tau > 0$. In addition, if we try to calculate $\langle \xi(t)\xi(t+\tau)\rangle$, we obtain:

$$\begin{aligned} g(0) = \langle \xi(t)\xi(t+\tau)\rangle &= \lim_{\Delta t\to 0}\left\langle \frac{\Delta W}{\Delta t}\frac{\Delta W}{\Delta t}\right\rangle \\ &= \lim_{\Delta t\to 0}\left\langle \frac{(\Delta W)^2}{(\Delta t)^2}\right\rangle = \lim_{\Delta t\to 0}\frac{1}{\Delta t} = \infty \end{aligned} \tag{4.9}$$

A function which has this property is the delta-function $\delta(\tau)$. Let us see what happens if we assume that ξ(t) is a noise source with the autocorrelation function:

$$\frac{dW}{dt} = \langle \xi(t)\xi(t+\tau)\rangle = \delta(\tau) \tag{4.10}$$

We now try using this assumption to solve the equation:

$$dx = \sigma dW \tag{4.11}$$

The solution to 4.11 is just $\mathrm{x(t)} = \sigma \mathrm{W(t)}$. If ξ(t) exists then we can write the

stochastic equation as $dW = \xi(\mathrm{t})dt$:

$$\begin{aligned} dx &= \sigma\xi(t)dt \\ \frac{dx}{dt} &= \sigma\xi(t) \\ x(t) &= \sigma\int_0^t \xi(s)ds \end{aligned} \tag{4.12}$$

Now, let us calculate the variance of x(t). This is:

$$\mathbb{V}(x(t)) = \left\langle x(t)^2 \right\rangle - \langle x(t)\rangle^2$$

Where $\langle x(t)\rangle = 0$ as $\xi(\mathrm{t})$ is a zero-mean process. Then, from equation 4.12

$$\begin{aligned} V(x(t)) &= \left\langle \sigma^2 \int_0^t \delta(s)ds \int_0^t \delta(v)dv \right\rangle \\ &= \sigma^2 \int_0^t \int_0^t \langle\delta(s)\delta(v)\rangle dsdv \\ &= \sigma^2 \int_0^t \int_0^t \delta(s-v)dsdv = \sigma^2 \int_0^t dv = \sigma^2 t \end{aligned} \tag{4.13}$$

which is the correct answer. Also, we can calculate the two-time auto-correlation function of $x(t)$. This is,

$$\begin{aligned} \langle x(t)x(t+\tau)\rangle &= \left\langle \sigma^2 \int_0^t \delta(s)ds \int_0^{t+\tau} \delta(v)dv \right\rangle = \sigma^2 \int_0^t \int_0^{t+\tau} \langle\delta(s-v)\rangle dsdv \\ &= \sigma^2 \int_0^t dv = \sigma^2 t \end{aligned} \tag{4.14}$$

which is also correct. So we have been able to obtain the correct solution to the stochastic differential equation by assuming that:

$$\xi(t) \equiv \frac{dW(t)}{dt}$$

exists and has a delta auto-correlation function.
This technique will work for any SDE that has purely additive noise, but it does not work for SDEs in which the noise dW multiplies a function of any of the variables. Thus one cannot use it to solve the equation $\mathbf{dx = xdW}$, for example.

Now, since the power spectrum of a process is the Fourier transform of the auto-correlation function, this means that the spectrum of ξ is:

$$S(v) = \int_{-\infty}^{\infty} \delta(t) e^{-2\pi v t} dt = 1 \tag{4.15}$$

This spectrum is the same for all values of the frequency v. This means that ξ(t) contains arbitrarily high frequencies, and thus infinitely rapid fluctuations. It also means that the power in ξ(t) is infinite. Both of these are further reflections of the fact that ξ(t) is an idealization, and cannot be truly realized by any real noise source. Because the spectrum contains equal amounts of all frequencies, it is referred to as "white" noise.

5 Brownian Motion:

5.1 Introduction to Brownian Motion

With a simple microscope, in 1827 Robert Brown observed that pollen grains in water move in haphazard manner. From a Newtonian perspective, this is surprising as force is required to initiate motion and cause changes in direction. Where does this force come from; could it be that the observed particles are in some sense active and 'alive,' generating their own motion?

The classic 1905 paper by Albert Einstein demonstrates that no active mechanism is necessary, and that the random forces generated by the thermally excited water molecules can account for the motion of the grains. This explanation was confirmed by Jean Perrin in 1908, for which he was awarded the Nobel prize in 1926.

Let us for simplicity indicate the position of the particle by a one-dimensional coordinate x (e.g. its vertical position); extension to more coordinates is trivial. According to Newtonian dynamics, the particle accelerates in response to forces it experience. When the particle at x is immersed in fluid, this includes in addition to external potential forces (e.g. due to gravity), a frictional force due to the fluid viscosity. The deterministic equation governing motion is them:

$$m\ddot{x} = -\frac{\partial V}{\partial x} - \frac{1}{\mu}\dot{x} \tag{5.1}$$

For a sphere of radius a, the viscous drag (and corresponding mobility μ) is given by:

$$\mu = \frac{1}{6\pi a \eta} \tag{5.2}$$

Where η is the specific viscosity of the fluid.
It is important to have a measure of the relative importance of the inertial and viscous terms in the above equation. Let us consider an object (not necessarily a solid sphere) of typical size a and density ρ, moving with velocity v in a fluid. The inertial force necessary to bring to rapidly change its velocity, e.g. to bring it to rest over a distance of the order of its size, is $F_{\text{inertial}} \sim mv(v/a) \sim \rho \mathrm{a}^2 v^2$. The dissipative force due to the

fluid viscosity is of order $F_{viscous} \sim \eta$av. The relative importance of the two forces is captured by the Reynolds number:

$$Re = \frac{F_{\text{inertial}}}{F_{\text{viscous}}} = \frac{\rho a v}{\eta} \tag{5.3}$$

Our physical experiences of motion in fluids relate to the realm of large Reynolds number: We are mostly interested in water and room temperature, which has a kinematic viscosity of $\frac{\eta}{\rho} \approx 10^{-6}$ m^2 s^{-1}; and for an animal swimming in water Re $\approx$ 1 mx1 ms$^{-1}/10^{-6}$ m^2 s$^{-1} = 10^6 \gg 1$ Even if the motive force is stopped, the animal will continue to move in the fluid due to its inertia. By contrast, cell and subcellular motion belong to the realm of low Reynolds numbers. For example, a typical bacterium is a few microns is size, and moves at velocities of around 30 μs^{-1}, translating to a Reynolds number of around $10^{-4} \ll 1$. For molecular motors, relevant length scales are of the order of 10 nm , with velocities of order of $1\mu s^{-1}$, leading to even smaller Re $\approx 10^{-8}$. The classic paper "Life at Low Reynolds Number" by Purcell contains many interesting observations about this limit.

At such small Reynolds numbers we can neglect the left-hand side of Eq. (1), concluding that velocity is proportional to external force:

$$m\dot{x} = -\frac{\partial V}{\partial x} - \frac{1}{\mu}\dot{x} \tag{5.4}$$

Of course, by the time we get down to the short scales of microns and below, we should no longer treat water as a continuous fluid; rather, it's particulate nature comes into play. The water molecules are constantly moving due to thermal fluctuations, and their impacts on the larger immersed objects results in a random force η(t), leading to the stochastic equation of motion

$$\dot{x} = -\mu\frac{\partial V}{\partial x} + \eta(t) \tag{5.5}$$

The stochastic (5.5) is the Langevin equation for the coordinate x. Different realizations of the stochastic force $\eta(t)$ lead to different values of x(t). The random impacts from all around an immersed object should average to zero over time, but there will be instantaneous fluctuations. We expect the random forces experienced over times longer than typical intervals between collisions to be uncorrelated, leading to:

$$\begin{aligned} \langle \eta(t) \rangle &= 0 \\ \langle \eta(t)\eta\left(t'\right) \rangle &= 2D\delta\left(t - t'\right) \end{aligned} \tag{5.6}$$

Where D is the measure of the measure of the strength of the fluctuating force.

5.2 Solving Brownian Equation from Newton 2^{nd} Law:

From fluid dynamics, we know that the pollen grain will experience friction from the liquid. This friction force is proportional to the negative of the momentum of the grain, so that:

$$F_{\text{friction}} = -\gamma \text{p} = -\gamma m v \tag{5.7}$$

Where m is the mass of the grain and v is its velocity. The constant of proportionality, γ, is usually referred to as the damping rate, and is given by $\gamma = 6\pi\frac{na}{m}$ Here a is the diameter of the pollen grain (assumed spherical), and η is the viscosity of the liquid.

We will only treat the motion in one dimension, since we lose nothing important by this restriction, and the extension to three dimensions is simple. The equation of motion for the position of the pollen grain is therefore:

$$m\frac{d^2x}{dt} = F_{\text{friction}} + F_{\text{fluct}}$$

$$m\frac{d^2x}{dt} = -\gamma p + F_{\text{fluct}} \tag{5.8}$$

Let us now consider equation 5.5. Realizing that the $\xi(t)$ associated with white noise has the same autocorrelation function, namely,

$$\langle \xi(t)\xi(t+\tau)\rangle = \delta(\tau)$$

We introduce the following term to represent our fluctuation F_{fluc} :

$$m\frac{d^2x}{dt} = -\gamma p + \sigma\xi(t) \tag{5.9}$$

Where σ controls the strength of fluctuation.
We know:

$$\frac{dx}{dt} = \frac{p}{m}, \tag{5.10}$$

$$\frac{d^2x}{dt^2} = \frac{dp}{dt} = -\gamma p + \sigma\xi(t). \tag{5.11}$$

Writing these equations in vector form, we have:

$$\frac{d}{dt}\begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{m} \\ 0 & -\gamma \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} + \begin{pmatrix} 0 \\ \sigma\xi(t) \end{pmatrix} \tag{5.12}$$

From $\frac{dW}{dt} = \xi(t)$,

$$\frac{d}{dt}\begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{m} \\ 0 & -\gamma \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} + \begin{pmatrix} 0 \\ \sigma \end{pmatrix}\frac{dW}{dt}$$
$$\begin{pmatrix} dx \\ dp \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{m} \\ 0 & -\gamma \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix}dt + \begin{pmatrix} 0 \\ \sigma \end{pmatrix}dW \tag{5.13}$$

To obtain the solution for the system of equations above, we note that dx depends upon p but dp does not depend on x . Let us first solve the equation for p , which assumes the form,

$$dp = -\gamma p dt + \sigma dW \tag{5.14}$$

This equation represents Ornstein-Uhlenbeck process as characterized by equation(3.27). We already solved this system to find the expectation and variance. Thus the solution becomes:

$$p(t) = e^{-\theta(t-S)}p_s + \sigma\int_S^t e^{-\gamma(t-T)}dW_t$$

Where for $S = 0$, then:

$$p(t) = e^{-\gamma t}p(0) + \sigma\int_0^t e^{-\gamma(t-T)}dW_t \tag{5.15}$$

The Expectation and variance become:

$$\mathbb{E}\left[p_t\right] = e^{-\gamma t}p_s,$$
$$\mathbb{V}\left(p_t\right) = \frac{\sigma^2}{2\gamma}\left(1 - e^{-2\gamma t}\right). \tag{5.16}$$

From statistical mechanics, we know that the average energy of the pollen grain in steady state must equal to the following:

$$\langle E\rangle = \left\langle p^2/2m\right\rangle = \frac{k_B T}{2} \tag{5.17}$$

For this to be satisfied, we require that $p(t)$ in equation (5.15) be independent of time as $t \to \infty$. Thus, we see that,

$$\lim_{t\to\infty}\mathbb{V}(p(t)) = \frac{\sigma^2}{2\gamma} = \mathbb{V}(p(t))_{SS} \tag{5.18}$$

Where the subscript SS indicates steady state.
Since the mean value of $p(t)$ tends to zero as $t \to \infty$, in the steady-state the mean is zero. This means from,

$$\mathbb{V}(p(t)) = \left\langle p^2 \right\rangle - \langle p \rangle^2$$
$$\mathbb{V}(p(t)) = \left\langle p^2 \right\rangle \tag{5.19}$$

Since $\langle p \rangle = 0$
. For consistently with statistical mechanics, we therefore require,

$$\langle E \rangle = \left\langle p^2/2m \right\rangle = \frac{\mathbb{V}(p)}{2m} = \frac{\sigma^2}{4\gamma m} = \frac{\sigma^2}{24\pi\eta d} \tag{5.20}$$

This tells us that the strength of the noise must be,

$$g = \sqrt{12\pi\eta k T} \tag{5.21}$$

We now turn to our second equation $\frac{dx}{dt} = \frac{p}{m}$. Resolving it is easy,

$$x(t) = \frac{1}{m}\int_0^t p(s)ds$$

From equation 5.15,

$$x(t) = \frac{1}{m}\int_0^t \left[e^{-\gamma s}p(0) + \sigma\int_0^t e^{-\gamma(s-s')}dW\left(s'\right)\right]ds$$
$$x(t) = \frac{p(0)}{m}\int_0^t e^{-\gamma s}ds + \frac{\sigma}{m}\int_0^t \left[\int_0^s e^{-\gamma(s-s')}dW\left(s'\right)\right]ds$$
$$x(t) = \frac{p(0)}{m}\int_0^t e^{-\gamma s}ds + \frac{\sigma}{m}\int_0^t \left[\int_0^s e^{-\gamma(s-s')}ds'\right]dW(s)$$
$$x(t) = \frac{1}{m\gamma}\left(1 - e^{-\gamma t}\right)p(0) + \frac{\sigma}{m\gamma}\int_0^t \left(1 - e^{-\gamma s}\right)dW(s) \tag{5.22}$$

The above expression, 5.22, is the complete solution for x(t). We see from this that the probability for x(t) is a Gaussian. We can now easily calculate the mean and variance, which are,

$$\langle x(t) \rangle = \left\langle \frac{1}{m\gamma}\left(1 - e^{-\gamma t}\right)p(0)\right\rangle + \left\langle \frac{\sigma}{m\gamma}\int_0^t \left(1 - e^{-\gamma s}\right)dW(s)\right\rangle$$

For a Weiner process, we know that $\langle W \rangle = 0$. Thus,

$$\langle x(t) \rangle = \left\langle \frac{1}{m\gamma}\left(1 - e^{-\gamma t}\right)p(0)\right\rangle$$

All of which are constants and thus,

$$\langle x(t) \rangle = \frac{1}{m\gamma}\left(1 - e^{-\gamma t}\right)p(0) \tag{5.23}$$

Now let us consider variance.

$$\mathbb{V}(x(t)) = \mathbb{V}\left(\frac{1}{m\gamma}\left(1-e^{-\gamma t}\right)p(0)\right) + \mathbb{V}\left(\frac{\sigma}{m\gamma}\int_0^t \left(1-e^{-\gamma s}\right)dW(s)\right)$$

Since $\frac{1}{m\gamma}\left(1-e^{-\gamma t}\right)p(0)$ is deterministic, its variance reduces to zero. For the left most term, we apply Ito's rule,

$$\mathbb{V}(x(t)) = \frac{\sigma^2}{(m\gamma)^2}\int_0^t \left(1-e^{-\gamma s}\right)^2 ds = \frac{\sigma^2 t}{(m\gamma)^2} + \frac{\sigma^2}{2m^2\gamma^3}\left[4e^{-\gamma t} - e^{-2\gamma t} - 3\right]$$

Langevin realized that the damping (or decay) rate γ was very fast, much faster than the time-resolution with which the particle could be observed in a real experiment (at least back in 1908). So this means that $\gamma t \gg 1$. In this case the variance of x(t) simplifies to the approximate expression:

$$\mathbb{V}(x(t)) \approx \frac{\sigma^2 t}{(m\gamma)^2} + \frac{\sigma^2}{2m^2\gamma^3} = \frac{\sigma^2}{(m\gamma)^2}\left(t - \frac{3}{2\gamma}\right) \approx \frac{\sigma^2 t}{(m\gamma)^2} = \left(\frac{kT}{3\pi\eta d}\right)t \qquad (5.24)$$

Thus, for times at which $t\gamma \gg 1$, the variance of the position of the particle will be proportional to time. This is exactly the same behavior as Wiener noise, and because of this Wiener noise is often referred to as Brownian motion. To test if our model for Brownian motion is accurate, one can perform an experiment in which the motion of a pollen grain is observed over time. If we record the amount by which the position changes in a fixed time interval T, and average the square of this over many repetitions of the experiment, then we can calculate the variance V (x(T)). This experiment was performed in 1910 by Smoluchouski and others, confirming the scaling of the variance with time. Since that time manv more experiments have been performed, and deviations from Wiener noise have been observed at very short time-scales

So how is it that x(t) can act just like the Wiener process? For this to be the case, it needs to be true that:

$$dx = \alpha dW, \quad \text{or equivalently} \quad \frac{dx}{dt} \approx \xi(t) \rightarrow (26)$$

for some constant α. Since $\frac{dx}{dt} = \frac{p}{m}$, this means that p(t) is effectively acting like $\xi(t)$. The reason for this is that the auto-correlation function of $p(t)$ decays at the rate γ, so that if γ is large then the auto-correlation function is sharply peaked, and approximates the auto-correlation function of ξ(t), which is a delta-function. Of course, when we say that γ is large, we actually mean that it is large compared to something else. In this case the something else is the time resolution (or time-scale), δt, over which the process is observed. By timescale we mean the time that elapses between observations of the process. In this case the process is $x(t)$. If we know the value of x(t) at time t = 0, and then again at a later time t + δt, then we cannot distinguish it from theWiener process so long as $\delta t \gg 1/\gamma$. So when we say that γ is large enough

that we cannot distinguish p(t) from ξ(t), we mean that γ is much larger than $1/\delta$t, where δt is the time resolution of our knowledge of p(t) and x(t).

5.3 White Noise and Brownian Motion

Brownian motion is the path taken by tiny particles in a viscous fluid due to being bombarded by the random thermal motion of the fluid molecules. There are two main modeling approaches. Einstein used a limited derivation of the Fokker-Plank equation to show that an ensemble of such particles obeys the diffusion equation. Langevin took a noise approach and showed that a particle with a small amount of momentum driven by small uncorrelated impacts follows a path with an exponentially decaying autocorrelation.

Turning back to white noise (I'll tie it all together eventually), assuming spatial and temporal homogeneity (the conditions in the infinite beaker are the same everywhere and do not change in time), then the tiny impacts to the particle constitute a sort of noise signal in time. Maybe they are the readings on an impact meter strapped to the particle. Because these impacts are 1) uncorrelated, 2) independent, and 3) comprised of an enormous amount of other collisions taking place in the fluid between hitting the Brownian particles, their magnitude has a Gaussian distribution.

If you take the Fourier transform (in real life, the FFT) of the impact signals for a large ensemble of Brownian particles and average them, you find that the power spectrum is constant over all frequencies, and that the power for a given frequency is distributed across the ensemble as a Gaussian distribution around the mean. Thus the impact signals are (on average) a combination of equal portions of all frequencies, we call it "white" noise, as in light comprised of all frequencies of visible light.

Back to the particles, their motion is the sum of a large number of these white noise impacts. If you are willing to consider this an integral in time, then you know that the power spectrum of this time integral will be proportional to $\frac{1}{f^2}$

This is the power spectrum of Brownian motion. Taking this the opposite way, from the Langevin equation you can see that the motion of a Brownian particle has an exponentially decaying autocorrelation. This corresponds to a power spectrum that decays as $\frac{1}{f^2}$. Calculating the derivative in frequency space, the derivative if Brownian motion looks like white noise.

None of this is either physically nor mathematically rigorous. But this is the general modeling approach used in physics and digital signal processing. However, these dirty models are the basis, and I might even say raison d'etre for the rigorous mathematical models. However, being dirty, there are probably multiple ways the white noise and Brownian motion can be defined. So it may simply be that when you read book A or paper X you need to use their definition.

6 Fokker-Planck Equation:

6.1 Deriving the Fokker-Planck Equation

Given a stochastic process x(t) with the Ito differential equation

$$dx = f(x,t)dt + g(x,t)dW \tag{6.1}$$

then the Fokker-Planck equation can be derived very easily. To do this we first calculate the differential equation for the mean value of an arbitrary function h(x).

Firstly, applying Ito's rule for h(x),

$$dh(t,\boldsymbol{x}) = \left(\frac{\partial h}{\partial x}\right) d\boldsymbol{x} + \left(\frac{dh}{\partial t}\right) dt + \frac{1}{2}\left(\frac{d^2h}{dx^2}\right)(d\boldsymbol{x})^2$$

The function $h(x)$ does not depend upon time,

$$dh(t,\boldsymbol{x}) = \left(\frac{\partial h}{\partial x}\right) d\boldsymbol{x} + \frac{1}{2}\left(\frac{d^2h}{d\boldsymbol{x}^2}\right)(d\boldsymbol{x})^2$$

Plugging equation 6.1 inside,

$$dh(t,x) = \left(\frac{\partial h}{\partial x}\right)[f(x,t)dt + g(x,t)dW] + \frac{1}{2}\left(\frac{d^2h}{dx^2}\right)(f(x,t)dt + g(x,t)dW)^2$$

The cross terms would not survive and using Ito's relation $dW^2 = \text{dt}$, we would have:

$$dh(t,\boldsymbol{x}) = \left(\frac{\partial h}{\partial x}\right)[f(x,t)dt + g(x,t)dW] + \frac{1}{2}\left(\frac{d^2h}{dx^2}\right)g^2(x,t)dt$$
$$dh(t,\boldsymbol{x}) = \left(\frac{\partial h}{\partial x}\right)f(x,t)dt + \frac{1}{2}\left(\frac{d^2h}{dx^2}\right)g^2(x,t)dt + \left(\frac{\partial h}{\partial x}\right)g(x,t)dW$$

Taking average on both sides, and recalling that $\langle W\rangle = 0$, we have:

$$d\langle h\rangle = \left\langle \left(\frac{\partial h}{\partial x}\right)f(x,t)dt \right\rangle + \frac{1}{2}\left\langle \left(\frac{d^2h}{dx^2}\right)g^2(x,t)dt \right\rangle$$
$$\frac{d\langle h\rangle}{dt} = \left\langle \left(\frac{\partial h}{\partial x}\right)f(x,t) \right\rangle + \frac{1}{2}\left\langle \left(\frac{d^2h}{dx^2}\right)g^2(x,t)dt \right\rangle$$
$$\frac{d\langle h\rangle}{dt} = \int_{-\infty}^{\infty}\left[f(x,t)\left(\frac{\partial h}{\partial x}\right) + \frac{1}{2}g^2(x,t)\left(\frac{d^2h}{dx^2}\right)\right]P(x,t)dx \tag{6.2}$$

We now perform integration by parts for both of the terms. Once for the first one and twice for the latter. Consider first the term $\int_{-\infty}^{\infty} f(x,t)\left(\frac{\partial h}{\partial x}\right)P(x,t)dx$.

We have:

$$u = P(x,t), \quad dv = \frac{dh}{dx}dx.$$

Then,

$$du = \frac{\partial P(x,t)}{\partial x}dx, \quad v = h(x).$$

Integration by parts yields:

$$\int_{-\infty}^{\infty} f(x,t)\left(\frac{\partial h}{\partial x}\right)P(x,t)dx = [h(x)f(x,t)P(x,t)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} h(x)\frac{\partial}{\partial x}(f(x,t)P(x,t))dx$$

Note that a valid PMF $P(x,t)$ must converge to zero at infinity. Thus, the first term in above expression should be zero and it should simplify to the following:

$$\int_{-\infty}^{\infty} f(x,t)\left(\frac{\partial h}{\partial x}\right)P(x,t)dx = -\int_{-\infty}^{\infty} h(x)\frac{\partial}{\partial x}(f(x,t)P(x,t))dx \tag{6.3}$$

Let us now consider the second term $\int_{-\infty}^{\infty} g^2(x,t)\left(\frac{d^2h}{dx^2}\right)P(x,t)dx$. We perform integration by parts twice on this. We have:

$$u = P(x,t)g^2(x,t), \quad dv = \frac{d^2h}{dx^2}.$$

Then,

$$du = \frac{\partial}{\partial x}\left[P(x,t)g^2(x,t)\right], \quad v = \frac{dh}{dx}.$$

Applying the integration by parts formula:

$$\int_{-\infty}^{\infty} g^2(x,t)\left(\frac{d^2h}{dx^2}\right)P(x,t)dx = -\int_{-\infty}^{\infty}\frac{\partial}{\partial x}\left[P(x,t)g^2(x,t)\right]\frac{dh}{dx}dx \tag{6.4}$$

Where the boundary term $\left[P(x,t)g^2(x,t)\frac{dh}{dx}\right]_{-\infty}^{\infty}$ once again goes to 0 .
Applying integration on 6.4 again.
We have:

$$u = \frac{\partial}{\partial x}\left[P(x,t)g^2(x,t)\right], \quad dv = \frac{dh}{dx}.$$

Then,

$$du = \frac{\partial^2}{\partial x^2}\left[P(x,t)g^2(x,t)\right], \quad v = h(x).$$

Applying the integration by parts formula:

$$\int_{-\infty}^{\infty} g^2(x,t)\left(\frac{d^2h}{dx^2}\right)P(x,t)dx = \int_{-\infty}^{\infty} h(x)\frac{\partial^2}{\partial x^2}\left[P(x,t)g^2(x,t)\right]dx \tag{6.5}$$

Where the boundary term $\left[h(x)\frac{\partial}{\partial x}\left[P(x,t)g^2(x,t)\right]\right]_{-\infty}^{\infty}$ once again goes to 0. In terms of 6.(5) and 6.(3), the original equation 6.(2) becomes:

$$\frac{d\langle h\rangle}{dt} = \int_{-\infty}^{\infty} -h(x)\frac{\partial}{\partial x}\left(f(x,t)P(x,t)\right)dx + \int_{-\infty}^{\infty}\frac{1}{2}h(x)\frac{\partial^2}{\partial x^2}\left[P(x,t)g^2(x,t)\right]dx$$

$$\frac{d\langle h\rangle}{dt} = \int_{-\infty}^{\infty} h(x)\left\{-\frac{\partial}{\partial x}\left(f(x,t)P(x,t)\right) + \frac{\partial^2}{\partial x^2}\left[P(x,t)g^2(x,t)\right]\right\}dx \tag{6.6}$$

We also know that the mean of f is given by,

$$\langle h\rangle = \int_{-\infty}^{\infty} h(x)P(x,t)dx \tag{6.7}$$

Thus, the derivative of mean can be written as,

$$\frac{d\langle h\rangle}{dt} = \frac{d}{dt}\int_{-\infty}^{\infty} h(x)P(x,t)dx = \int_{-\infty}^{\infty} h(x)\frac{\partial}{\partial t}P(x,t)dx \tag{6.8}$$

Equation 6.8 and 6.(6) together,

$$\int_{-\infty}^{\infty} h(x)\frac{\partial}{\partial t}P(x,t)dx = \int_{-\infty}^{\infty} h(x)\Bigg\{-\frac{\partial}{\partial x}\left[f(x,t)P(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[P(x,t)g^2(x,t)\right]\Bigg\}dx. \tag{6.9}$$

Equation 6.9 should hold for any h(x). Thus, the Fokker-Plank Equation becomes the following,

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}(f(x,t)P(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[D^2(x,t)P(x,t)\right] \tag{6.10}$$

Where we have defined $D^2(x,t) = g^2(x,t)$.
We can write equation 6.10 as a probability current:

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}\left[f(x,t)P(x,t) + \frac{1}{2}\frac{\partial}{\partial x}\left(D^2(x,t)P(x,t)\right)\right]$$

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}J(x,t). \tag{6.11}$$

Where

$$J(x,t) = f(x,t)P(x,t) + \frac{1}{2}\frac{\partial}{\partial x}\left(D^2(x,t)P(x,t)\right).$$

The relation between P and J , as given by 6.11 implies:

- J(x) is the probability current
- $J(x,t)$ is the rate at which probability is flowing across the point x at time t.

To see this, consider the probability that x lies within the narrow interval $[a, a+x]$. This probability is approximately P(a, t)x. Now consider the rate of change of this probability as shown in the figure below:

ADD FIGURE HERE

This rate of change is given by the difference between the rate at which probability is flowing into the interval from the left, and the rate that it is flowing out from the right. Denoting the rate of flow of probability across the point x at time t as $R(x,t)$, we have:

$$\frac{\partial}{\partial t}[P(a,t)\Delta x] = R(a,t) - R(a+\Delta x, t) \tag{6.12}$$

Dividing both sides by Δx, and taking the limit as $\Delta x \to 0$, we get:

$$\frac{\partial}{\partial t}P(a,t) = -\lim_{\Delta x \to 0}\frac{R(a+\Delta x,t) - R(a,t)}{\Delta x} = -\frac{\partial}{\partial a}R(a,t) \tag{6.13}$$

Comparing this with equation (6.11), we see that J(x, t) is indeed the rate of flow of the probability across the point x .

Boundary Conditions: Absorbing and Reflecting Boundaries:

To solve an FP equation, one may also need to specify the boundary conditions. If x takes values on the entire real line, then this is unnecessary since we know that P tends to zero as $x \to \pm\infty$, and this will be reflected in the initial condition, being the choice for P(x, t) at t = 0. However, if x has some finite domain, say the interval $[a, b]$, then we need to specify what happens at the boundaries a and b . The three most common possibilities are as follows.

1. **Absorbing boundaries**. An absorbing boundary is one in which the particle is removed immediately hits the boundary. This means that the probability that particle is on the boundary is always zero, and this situation is therefore described by the condition

$$\mathrm{P(c,t)} = 0 \tag{6.14}$$

where c is the location of the absorbing boundary.

2. **Reflecting boundaries**. A reflecting boundary is one for which the particle cannot pass through. This means that the probability current must be zero across the boundary, and is therefore given by the condition

$$J(c,t) = 0 \tag{6.15}$$

where c is the location of the reflecting boundary.

3. **Periodic boundaries**. In this case the two ends (boundaries) of the interval are connected together. This means that the particle is moving on a closed loop such as a circle in one dimension, or a torus in two dimensions. In this case, since the two ends describe the same physical location, both the probability density

and the probability current must be the same at both ends. This is therefore described by the two conditions

$$\mathrm{P(a,t) = P(b,t),} \tag{6.16}$$

$$\mathrm{J(a,t) = J(b,t).} \tag{6.17}$$

where the interval in which the particle moves is [a, b].

These three kinds of boundary conditions can also be applied to FP equations in more than one dimension. For reflecting boundaries this means setting to zero the dot product of the vector current with the vector normal to the surface of the boundary.

6.2 Kolmogorov Backward Equation:

Consider an SDE of the form:

$$d\boldsymbol{x} = \mu\left(X_t, t\right) dt + \sigma\left(X_t, t\right) d\boldsymbol{w}$$

We can define the associated Fokker Plank equation for the above,

$$\frac{\partial}{\partial t} p(x,t) = -\frac{\partial}{\partial x_t}\left(\mu\left(x_t, t\right) p(x,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x_t^2}\left[\sigma^2(x,t) p(x,t)\right]$$

Where $P(x,t)$ is the probability current. We can think of Fokker Plank Equation as a normal distribution being transformed into an arbitrary complex distribution according to the drift and diffusion parameters $\mu\left(x_t\right)$ and $\sigma\left(x_t\right)$. In fact, if we think in terms of conditional probability densities just like we did way back when considering the conditional distribution to find auto-correlation function, then Fokker Plank Equation basically tells us the evolution:

$$p\left(x, t \mid x_0, t_0\right) \tag{6.18}$$

What the Kolmogorov backward Equation allows us answer is how the probability of x_0 at a later point in time changes as we change x_t at an earlier point in time. In completeness, it answers:
'How does the probability of $\boldsymbol{x_0}$ at the later point in time s change, as we slowly evolve the probability distribution backwards through time and condition on $\boldsymbol{x_t}$ "

It is defined as following:

$$-\frac{\partial}{\partial t} p\left(x_0 \mid x_t\right) = \mu\left(x_t\right) \frac{\partial}{\partial x_0} p\left(x_0 \mid x_t\right) + \frac{1}{2}\sigma^2\left(x_t\right) \frac{\partial^2}{\partial x_0} p\left(x_0 \mid x_t\right) \tag{6.19}$$

To derive this, we are going to consider the probability distribution $p\left(x, t \mid x_0, t_0\right)$ which describes the system at position at x at time t , given that it was at position

x_0 at time t_0. Throughout our derivation, we would assume without proof that the probability distribution $p\left(x,t \mid x_0,t_0\right)$ is a continuous function whose derivatives (to arbitrary order) are finite. (This ofcourse is hand-wavy and more formal arguments must demonstrate that this assumption holds.)

Before we begin, we first establish an important equation that $P\left(x,t \mid x_0,t_0\right)$ must satisfy. Consider the example of fish movement, and imagine a fish that starts in location x_0 at time t_0 and ends up at location x at time t . We can obtain an expression for $P\left(x,t \mid x_0,t_0\right)$ by considering all of the potential locations, x_1, of the fish at some intermediate point in time, t_1. In particular, $P\left(x,t \mid x_0,t_0\right)$ can be expressed as the probability of moving from x_0 to x_1 between
times t_0 and t_1, and then moving from x_1 to x between times t_1 and t , evaluated over all possible intermediate states, x_1. We can write this logical statement mathematically as,

$$p\left(x,t \mid x_0,t_0\right) = \int p\left(x,t \mid x_1,t_1\right) p\left(x_1,t_1 \mid x_0,t_0\right) dx_1 \tag{6.20}$$

where this integral (and those that follow) is evaluated over the range of possible values of the random variable x_1. Equation 6.20 is known as the Chapman-Kolmogorov equation.

Let us now derive the backward equation (6.19). Consider an intermediate point in time, t_1, between the present time t and the initial time point t_0 but one that is very close to t_0. In this case we can write $t_1 = t_0 + \Delta t$ where Δt is very small. The crux of the derivation revolves around the assumption that the change in the random variable X(t) over the short time period Δt is small enough that we can use a Taylor series with respect to this change. In particular, after a small amount of time Δt elapses, the value of X is assumed to change by a small amount Δx. Thus we can write the value of X at time t_1 as $x_1 = x_0 + \Delta x$.

Our goal is to derive an expression for the derivative, $\frac{\partial P}{\partial t_0}$. We start by using the definition for the derivative: 1

$$\frac{\partial p\left(x,t \mid x_0,t_0\right)}{\partial t_0} = \lim_{\Delta t \to 0} \frac{p\left(x,t \mid x_0,t_0+\Delta t\right) - p\left(x,t \mid x_0,t_0\right)}{\Delta t} \tag{6.21}$$

To obtain the desired expression, we must obtain an expression for the ratio:

$$\frac{p\left(x,t \mid x_0,t_0+\Delta t\right) - p\left(x,t \mid x_0,t_0\right)}{\Delta t} \tag{6.22}$$

First, we can replace $p\left(x,t \mid x_0,t_0\right)$ in this ratio with the Chapman-Kolmogorov equation 6.20. In addition, we know that $\int p\left(x_1,t_1 \mid x_0,t_0\right) dx_1 = 1$ because the fish must be located somewhere at time t_1. Consequently, we are free to replace $p\left(x,t \mid x_0,t_0+\Delta t\right)$ in 6.21 with:

$$p\left(x,t \mid x_0,t_0+\Delta t\right) = p\left(x,t \mid x_0,t_0+\Delta t\right) \int p\left(x_1,t_1 \mid x_0,t_0\right) dx_1$$

Since $\int p\left(x_1,t_1 \mid x_0,t_0\right) dx_1$ is just equal to 1.

Making these adjustments into 6.22:

$$\frac{1}{\Delta t}\left[p\left(x,t\mid x_0,t_0+\Delta t\right)\int p\left(x_1,t_1\mid x_0,t_0\right)dx_1-\int p\left(x,t\mid x_1,t_1\right)p\left(x_1,t_1\mid x_0,t_0\right)dx_1\right]$$

Let $t_1 = t_0 + \Delta t$, then:

$$\frac{1}{\Delta t}\left[p\left(x,t\mid x_0,t_1\right)\int p\left(x_1,t_1\mid x_0,t_0\right)dx_1-\int p\left(x,t\mid x_1,t_1\right)p\left(x_1,t_1\mid x_0,t_0\right)dx_1\right]$$

Moving $p\left(x,t\mid x_0,t_1\right)$ inside the integral and taking $p\left(x_1,t_1\mid x_0,t_0\right)$ common:

$$\frac{1}{\Delta t}\int\left\{p\left(x,t\mid x_0,t_1\right)-p\left(x,t\mid x_1,t_1\right)\right\}p\left(x_1,t_1\mid x_0,t_0\right)dx_1$$

Next, we use the fact that $x_1 = x_0 + \Delta x$ and take the Taylor series of the term within the curly brackets:

$$\frac{1}{\Delta t}\int\left\{p\left(x,t\mid x_0,t_1\right)-p\left(x,t\mid x_0+\Delta x,t_1\right)\right\}p\left(x_1,t_1\mid x_0,t_0\right)dx_1$$

Taking the Taylor series,

$$\frac{1}{\Delta t}\int\left\{-\Delta x\frac{\partial p\left(x,t\mid x_0,t_1\right)}{\partial x_0}-\frac{\Delta x^2}{2}\frac{\partial^2 p\left(x,t\mid x_0,t_1\right)}{\partial x_0^2}-O\left(\Delta x^3\right)\right\}p\left(x_1,t_1\mid x_0,t_0\right)dx_1$$

At this point, we drop the higher-order terms $O\left(\Delta x^3\right)$. Doing so is equivalent to assuming that process does not make large "jumps" in small time intervals (see p. 327 in Allen 2003). Replacing Δx, with $x_1 - x_0$ and factoring out terms that do not depend on x_1 leaves:

$$\frac{1}{\Delta t}\int-\left(x_1-x_0\right)\frac{\partial p\left(x,t\mid x_0,t_1\right)}{\partial x_0}dx_1-\frac{1}{2\Delta t}\frac{\partial^2 p\left(x,t\mid x_0,t_1\right)}{\partial x_0^2}\int\left(x_1-x_0\right)^2\frac{\partial p\left(x,t\mid x_0,t_1\right)}{\partial x_0}dx_1$$

Finally, we put this in the limit by going back to equation 6.21:

$$\begin{aligned}\frac{\partial p\left(x,t\mid x_0,t_0\right)}{\partial t_0}=\lim_{\Delta t\to 0}\frac{1}{\Delta t}&\int-\left(x_1-x_0\right)\frac{\partial p\left(x,t\mid x_0,t_1\right)}{\partial x_0}dx_1\\&-\frac{1}{2\Delta t}\frac{\partial^2 p\left(x,t\mid x_0,t_1\right)}{\partial x_0^2}\int\left(x_1-x_0\right)^2\frac{\partial p\left(x,t\mid x_0,t_1\right)}{\partial x_0}dx_1.\end{aligned}\qquad(6.23)$$

We replace t_1 with $t_0 + \Delta t$ and take the limit as $\Delta t \to 0$, allowing us to write the above as:

$$\frac{\partial p\left(x,t \mid x_0,t_0\right)}{\partial t_0} = -\mu\left(x_0\right)\frac{\partial p\left(x,t \mid x_0,t_1\right)}{\partial x_0} - \frac{1}{2}\sigma\left(x_0\right)\frac{\partial^2 p\left(x,t \mid x_0,t_1\right)}{\partial x_0^2} \tag{6.24}$$

Where,

$$\mu\left(x_0\right) = \lim_{\Delta t\to 0}\frac{\int -\left(x_1 - x_0\right)p\left(x_1, t_0+\Delta t \mid x_0,t_0\right)dx_1}{\Delta t}, \tag{6.25}$$

$$\sigma^2\left(x_0\right) = \lim_{\Delta t\to 0}\frac{\int \left(x_1 - x_0\right)^2 p\left(x_1, t_0+\Delta t \mid x_0,t_0\right)dx_1}{\Delta t}. \tag{6.26}$$

Equations 6.25 and 6.26 can be more easily interpreted in terms of the expected rate of change. By definition, the expected value $E[g(X)]$ of a function $g(X)$ is given by multiplying $g(X)$ by the probability density function for the random variable X and integrating over all possible values of the random variable. Thus, the expected change in the random variable over a time step Δt, raised to the power k, is defined as:

$$\mathbb{E}\left[\left(X\left(t_0+\Delta t\right) - X\left(t_0\right)\right)^k \mid X\left(t_0\right) = x_0\right] = \int \left(x_1 - x_0\right)^k p\left(x_1, t_0+\Delta \mathrm{t} \mid x_0,t_0\right)dx_1 \tag{6.27}$$

6.3 Stationary Solution for one dimension:

When the FP equation is one dimensional (that is, has only a single variable, x), one can fairly easily calculate its stationary or steady-state solutions. A stationary solution is defined as one in which P(x, t) does not change with time. The stationary distribution represents the long-term behavior of the stochastic system described by the Fokker-Planck equation. It tells us how the probabilities of the system's states are distributed when the system has reached equilibrium. In physical systems, particularly in thermodynamics and statistical mechanics, the stationary distribution often corresponds to the thermodynamic equilibrium distribution. For example, in systems governed by Boltzmann statistics, the stationary distribution is the Boltzmann distribution, as we will show.

The differential equation that describes the stationary solutions is obtained by setting:

$$\frac{\partial P}{\partial t} = 0 \tag{6.28}$$

Thus, equation 6.11 becomes:

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}J(x,t)$$
$$0 = -\frac{\partial}{\partial x}J(x,t)$$
$$0 = -\frac{\partial}{\partial x}[f(x,t)P(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[D^2(x,t)P(x,t)\right]$$

If we have reflecting boundary conditions, then J = 0 on at least one boundary, and thus J = 0 everywhere.

$$\frac{d}{dx}\left[D^2(x,t)P(x,t)\right] = 2[f(x,t)P(x,t)] \tag{6.29}$$

Defining a new function $\xi(x) = D(x)P(x)$, we see that this equation is a just a linear differential equation for h :

$$\frac{d\xi}{dx} = \left[\frac{2f(x)}{D(x)}\right]\xi \tag{6.30}$$

The solution is,

$$P(x) = \frac{1}{\mathcal{N}D(x)}\exp\left[\int_a^x \frac{f(u)}{D(u)}du\right] \tag{6.31}$$

where the particle moves in the interval [a, b], and $\mathcal{N}$ is a constant chosen so that :

$$\mathcal{N} = \int_a^b P(x)dx = 1 \tag{6.32}$$

Notice that this is nothing but the Maxwellian-Boltzmann distribution from statistical mechanics. Thus, we recover our normal physics under equilibrium when $t \to \infty$.

If we have periodic boundary conditions, then J will not necessarily vanish. Neither is J a free parameter, however; as we will see, it is completely determined by the assumption of stationarity. In this case the equation for the stationary solution, P(x), is given by:

$$\frac{d}{dx}[D(x)P(x)] = 2f(x)P(x) - J \tag{6.33}$$

Once again defining $\xi(x) = D(x)P(x)$, this is a linear equation for ξ, but this with a constant driving term:

$$\frac{d\xi}{dx} = \left[\frac{2f(x)}{D(x)}\right]\xi - J \tag{6.34}$$

The solution is:

$$P(x) = \left[\frac{Z(x)}{D(x)}\right]\left\{P(a)\left[\frac{D(a)}{Z(a)}\right] - 2J\int_a^x \frac{du}{Z(u)}\right\}. \tag{6.35}$$

Where we have defined:

$$Z(x) = \exp\left[\int_a^x \frac{f(u)}{D(u)} du\right] \tag{6.36}$$

Now we apply the periodic boundary condition $P(a) = P(b)$ (note that we have already applied the boundary condition on J by making J constant) and this gives:

$$J = \frac{P(a)}{2\int_a^b \frac{du}{Z(u)}}\left[\frac{D(a)}{Z(a)} - \frac{D(b)}{Z(b)}\right] \tag{6.37}$$

The solution is therefore:

$$P(a) = P(a)\frac{\left[\frac{D(b)}{Z(b)}\int_a^x \frac{du}{Z(u)} - \frac{D(a)}{Z(a)}\int_x^b \frac{du}{Z(u)}\right]}{\frac{D(x)}{Z(x)}\int_a^b \frac{du}{Z(u)}} \tag{6.38}$$

7 Reverse-Time Stochastic Equation

7.1 Anderson's Formula

We know that the general SDE assumes the form:

$$dx = \mu dt + \sigma dw \tag{7.1}$$

We can define the associated Fokker Plank equation or the Kolmogorov Forward Equation as following:

$$\partial_t p(x,t) = -\partial x_t \left[\mu\left(x_t\right) p\left(x_t\right)\right] + \frac{1}{2}\partial_{x_t}^2\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right] \tag{7.2}$$

Where $P(x,t)$ is the probability current.
We can think of equation (2.2) as a normal distribution being transformed into an arbitrary complex distribution according to the drift and diffusion parameters $\mu\left(x_t\right)$ and $\sigma\left(x_t\right)$. The Kolmogorov Backward Equation gives the following:

$$-\partial_t p\left(x_s \mid x_t\right) = \mu\left(x_t\right)\partial_{x_t} p\left(x_s \mid x_t\right) + \frac{1}{2}\sigma^2\left(x_t\right)\partial_{x_t}^2 p\left(x_s \mid x_t\right) \tag{7.3}$$

Which answers the question:
'How does the probability of $\boldsymbol{x_s}$ at the later point in time s change, as we slowly evolve the probability distribution backwards through time and condition on $\boldsymbol{x_t}$ "

We would now like to find a reverse Fokker Plank equation. This basically offers a partial differential equation which we can solve backward in time, which would correspond to evolving the arbitrarily complex distribution backwards to our original Normal distribution.

This result was produced in the paper, **Reverse-time Diffusion Equation Models** in 1982 by BDO Anderson and is cited by Yang. Essentially, we are going to be using Bayes Theorem to make headways. In terms of x_s and x_t where x_s represents our

initial distribution and x_t represents our final distribution, we can write the marginal as following:

$$p\left(x_s, x_t\right) = p\left(x_s \mid x_t\right) p\left(x_t\right) \tag{7.4}$$

With the time ordering $t \leq s$.
Consider the Left-hand side. First multiplying both sides of Bayes theorem with minus one and taking the derivative with respect to time t via product rule, we obtain:

$$\begin{aligned} -\partial_t p\left(x_s, x_t\right) &= -\partial_t \left[p\left(x_s \mid x_t\right) p\left(x_t\right)\right] \\ &= -\partial_t p\left(x_s \mid x_t\right) p\left(x_t\right) - p\left(x_s \mid x_t\right) \partial_t p\left(x_t\right). \end{aligned} \tag{7.5}$$

Where $-\partial_t p\left(x_s \mid x_t\right)$ is the Kolmogorov Backward Equation and $\partial_t p\left(x_t\right)$ is the Kolmogorov Forward Equation. Plugging 7.2 and 7.3:

$$\begin{aligned} -\partial_t p(x_s, x_t) = &\left(\mu(x_t)\partial_x p(x_s \mid x_t) + \frac{1}{2}\sigma^2(x_t)\partial_x^2 p(x_s \mid x_t)\right) p(x_t) \\ &+ p(x_s \mid x_t)\left(\partial_x\left[\mu(x_t)p(x_t)\right] + \frac{1}{2}\partial_x^2\left[\sigma^2(x_t)p(x_t)\right]\right). \end{aligned} \tag{7.6}$$

The derivative occuring in the backward Kolmogorov equation are:

$$\begin{aligned} \partial_{x_t} p\left(x_s \mid x_t\right) &= \partial_{x_t}\left[\frac{p\left(x_s, x_t\right)}{p\left(x_t\right)}\right], \\ \partial_{x_t} p\left(x_s \mid x_t\right) &= \frac{\partial_{x_t} p\left(x_s, x_t\right) p\left(x_t\right) - p\left(x_s, x_t\right) \partial_{x_t} p\left(x_t\right)}{p^2\left(x_t\right)}, \\ \partial_{x_t} p\left(x_s \mid x_t\right) &= \frac{\partial_{x_t} p\left(x_s, x_t\right)}{p\left(x_t\right)} - \frac{p\left(x_s, x_t\right) \partial_{x_t} p\left(x_t\right)}{p^2\left(x_t\right)}. \end{aligned}$$

The next step is to evaluate the derivative of the products in the forward Kolmogorov equation.

$$\partial_{x_t}\left[\mu\left(x_t\right) p\left(x_t\right)\right] = \partial_{x_t}\mu\left(x_t\right) p\left(x_t\right) + \mu\left(x_t\right) \partial_{x_t} p\left(x_t\right), \tag{7.7}$$

$$\partial_{x_t}^2\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right] = \partial_{x_t}^2\sigma^2\left(x_t\right) p\left(x_t\right) + 2\partial_{x_t} p\left(x_t\right) \partial_{x_t}\sigma^2\left(x_t\right) + \sigma^2\left(x_t\right) \partial_{x_t}^2 p\left(x_t\right). \tag{7.8}$$

What follows is a long algebra gymnastics.Substituting the derivatives of the probability distribution accordingly we obtain:

$$-\partial_t p(x_s, x_t) = -\partial_t \left[p(x_s|x_t)p(x_t)\right]$$

$$-\partial_t p(x_s, x_t) = -\partial_t p(x_s|x_t)p(x_t) - p(x_s|x_t)\,\partial_t p(x_t)$$

$$\begin{aligned}-\partial_t p(x_s, x_t) = &\left(\mu(x_t)\partial_{x_t} p(x_s|x_t) + \frac{1}{2}\sigma^2(x_t)\partial^2_{x_t} p(x_s|x_t)\right)p(x_t) \\ &+ p(x_s|x_t)\left(\partial_{x_t}\left[\mu(x_t)p(x_t)\right] - \frac{1}{2}\partial^2_{x_t}\left[\sigma^2(x_t)p(x_t)\right]\right)\end{aligned}$$

Plugging equation (7.7)

$$\begin{aligned}-\partial_t p\left(x_s, x_t\right) = &\mu\left(x_t\right)\partial_{x_t} p\left(x_s \mid x_t\right) p\left(x_t\right) + \frac{1}{2}\sigma^2\left(x_t\right)\partial^2_{x_t} p\left(x_s \mid x_t\right) p\left(x_t\right) \\ &+ p\left(x_s \mid x_t\right)\partial_{x_t}\mu\left(x_t\right) p\left(x_t\right) + p\left(x_s \mid x_t\right)\mu\left(x_t\right)\partial_{x_t} p\left(x_t\right).\end{aligned}$$

Plugging equation (7.5):

$$\begin{aligned}-\partial_t p\left(x_s, x_t\right) = &\mu\left(x_t\right)\left[\frac{\partial_{x_t} p\left(x_s, x_t\right)}{p\left(x_t\right)} - \frac{p\left(x_s, x_t\right)\partial_{x_t} p\left(x_t\right)}{p^2\left(x_t\right)}\right] p\left(x_t\right) \\ &+ p\left(x_s \mid x_t\right)\partial_{x_t}\mu\left(x_t\right) p\left(x_t\right) + p\left(x_s \mid x_t\right)\mu\left(x_t\right)\partial_{x_t} p\left(x_t\right) \\ &+ \frac{1}{2}\sigma^2\left(x_t\right)\partial^2_{x_t} p\left(x_s \mid x_t\right) p\left(x_t\right) - \frac{1}{2} p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right].\end{aligned}$$

Using $p\left(x_s, x_t\right) = p\left(x_s \mid x_t\right) p\left(x_t\right)$:

$$\begin{aligned}-\partial_t p\left(x_s, x_t\right) = &\mu(x_t)\left[\partial_{x_t} p\left(x_s, x_t\right) - \underline{p\left(x_s \mid x_t\right)\partial_{x_t} p\left(x_t\right)}\right] \\ &+ p\left(x_s \mid x_t\right)\partial_{x_t}\mu(x_t)p(x_t) + p\left(x_s \mid x_t\right)\underline{\mu(x_t)\partial_{x_t} p(x_t)} \\ &+ \frac{1}{2}\sigma^2(x_t)\partial^2_{x_t} p\left(x_s \mid x_t\right) p(x_t) - \frac{1}{2} p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2(x_t)p(x_t)\right].\end{aligned}$$

$$\begin{aligned}-\partial_t p\left(x_s, x_t\right) = &\mu(x_t)\partial_{x_t} p(x_s, x_t) + p(x_s \mid x_t)\partial_{x_t}\mu(x_t)p(x_t) \\ &+ \frac{1}{2}\sigma^2(x_t)\partial^2_{x_t} p(x_s \mid x_t)p(x_t) - \frac{1}{2} p(x_s \mid x_t)\partial^2_{x_t}\left[\sigma^2(x_t)p(x_t)\right].\end{aligned}$$

Simplifying further, we notice the first two terms arise from a product rule:

$$-\partial_t p(x_s, x_t) = \partial_{x_t}\left[\mu(x_t)p(x_s, x_t)\right] + \frac{1}{2}\sigma^2(x_t)\partial^2_{x_t}p(x_s \mid x_t)p(x_t) - \frac{1}{2}p(x_s \mid x_t)\partial^2_{x_t}\left[\sigma^2(x_t)p(x_t)\right].$$

In order to transform the partial differential equation above into a form from which we can deduce an equivalent stochastic differential equation, we match the terms of the second order derivatives (given by the left-most expression in equation above with the following identity,

$$\begin{aligned}\frac{1}{2}\partial^2_{x_t}\left[p(x_s, x_t)\sigma^2(x_t)\right] &= \frac{1}{2}\partial^2_{x_t}\left[p(x_s \mid x_t)p(x_t)\sigma^2(x_t)\right] \\ &= \frac{1}{2}\partial^2_{x_t}p(x_s \mid x_t)p(x_t)\sigma^2(x_t) + \partial_{x_t}\left[p(x_t)\sigma^2(x_t)\right]\partial^2_{x_t}p(x_s \mid x_t) \\ &+ \frac{1}{2}p(x_s \mid x_t)\partial^2_{x_t}\left[p(x_t)\sigma^2(x_t)\right].\end{aligned}$$

by observing that the colored terms occur in both equations. We can see from the expansion of the derivative above that we can combine the terms in our derivation if we expand the "center term". Furthermore, we can employ the identity $-\frac{1}{2}X = -X + \frac{1}{2}X$ to obtain:

$$\begin{aligned}-\partial_t p(x_s, x_t) = \partial_{x_t}\left[\mu(x_t)p(x_s, x_t)\right] &+ \frac{1}{2}\sigma^2(x_t)\partial^2_{x_t}p(x_s \mid x_t)p(x_t) \\ &- \frac{1}{2}p(x_s \mid x_t)\partial^2_{x_t}\left[\sigma^2(x_t)p(x_t)\right] \pm \partial_{x_t}p(x_s \mid x_t)\partial_{x_t}\left[p(x_t)\sigma^2(x_t)\right].\end{aligned}$$

Consider the term $\frac{1}{2}p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2\left(x_t\right)p\left(x_t\right)\right]$. Applying the identity $-\frac{1}{2}X = -X + \frac{1}{2}X$,

$$\begin{aligned}-\partial_t p\left(x_s, x_t\right) = \partial_{x_t}\left[\mu\left(x_t\right)p\left(x_s, x_t\right)\right] &+ \frac{1}{2}\sigma^2\left(x_t\right)\partial^2_{x_t}p\left(x_s \mid x_t\right)p\left(x_t\right) \\ -p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2\left(x_t\right)p\left(x_t\right)\right] &+ \frac{1}{2}p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2\left(x_t\right)p\left(x_t\right)\right] \\ \pm\partial_{x_t}p\left(x_s \mid x_t\right)\partial_{x_t}&\left[p\left(x_t\right)\sigma^2\left(x_t\right)\right]\end{aligned}$$

Thus,

$$\begin{aligned}-\partial_t p\left(x_s, x_t\right) = \partial_{x_t}\left[\mu\left(x_t\right)p\left(x_s, x_t\right)\right] &+ \frac{1}{2}\sigma^2\left(x_t\right)\partial^2_{x_t}p\left(x_s \mid x_t\right)p\left(x_t\right) \\ +\frac{1}{2}p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2\left(x_t\right)p\left(x_t\right)\right] &- p\left(x_s \mid x_t\right)\partial^2_{x_t}\left[\sigma^2\left(x_t\right)p\left(x_t\right)\right] \\ -\partial_{x_t}p\left(x_s \mid x_t\right)\partial_{x_t}&\left[p\left(x_t\right)\sigma^2\left(x_t\right)\right].\end{aligned}$$

Where the red indicates a product rule term:

$$
\begin{aligned}
-\partial_t p\left(x_s, x_t\right) =& \partial_{x_t}\left[\mu\left(x_t\right) p\left(x_s, x_t\right)\right]+\frac{1}{2} \partial_{x_t}^2\left[p\left(x_s \mid x_t\right) \sigma^2\left(x_t\right)\right] \\
& -\partial_{x_t}\left[p\left(x_s \mid x_t\right) \sigma^2\left(x_t\right)\right]
\end{aligned}
$$

What remains to be done is to combine the joint probability and the conditional probability in the first order derivative terms to combine them,

$$
\begin{aligned}
-\partial_t p\left(x_s, x_t\right) =& \partial_{x_t}\left[\mu\left(x_t\right) p\left(x_s, x_t\right)-p\left(x_s \mid x_t\right) \partial_{x_t}\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right]\right] \\
& +\frac{1}{2} \partial_{x_t}^2\left[p\left(x_s \mid x_t\right) \sigma^2\left(x_t\right)\right] \\
-\partial_t p\left(x_s, x_t\right) =& \partial_{x_t}\left[p\left(x_s, x_t\right)\left(\mu\left(x_t\right)-\frac{1}{p\left(x_t\right)} \partial_{x_t}\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right]\right)\right]+\frac{1}{2} \partial_{x_t}^2\left[p\left(x_s \mid x_t\right) \sigma^2\left(x_t\right)\right] \\
-\partial_t p\left(x_s, x_t\right) =&-\partial_{x_t}\left[p\left(x_s, x_t\right)\left(-\mu\left(x_t\right)+\frac{1}{p\left(x_t\right)} \partial_{x_t}\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right]\right)\right]+ \\
& \frac{1}{2} \partial_{x_t}^2\left[p\left(x_s \mid x_t\right) \sigma^2\left(x_t\right)\right]
\end{aligned}
$$

the result of which is in the form of a Kolmogorov forward equation, although using the joint probability distribution $\mathrm{p}\left(x_s, x_t\right)$. For the time ordering of $t \leq s$, we can observe that the term $-\partial_t p\left(x_s, x_t\right)$ describes the change of the probability distribution as we move backward in time. In accordance with Leibniz' rule we can marginalize over x_s without interfering with the partial derivative ∂_t, to obtain:

$$
\begin{aligned}
-\partial_t p\left(x_t\right) =&-\partial_{x_t}\left[p\left(x_t\right)\left(-\mu\left(x_t\right)+\frac{1}{p\left(x_t\right)} \partial_{x_t}\left[\sigma^2\left(x_t\right) p\left(x_t\right)\right]\right)\right] \\
& +\frac{1}{2} \partial_{x_t}^2\left[p\left(x_t\right) \sigma^2\left(x_t\right)\right] .
\end{aligned}
$$

and introduce the time reversal $\tau=1-t$ which, with respect to the integration with respect to the flow of time, yields:

$$
\begin{aligned}
-\partial_t p\left(x_t\right) =&-\partial_\tau p\left(x_{1-\tau}\right) \\
=&-\partial_{x_t}\left[p\left(x_{1-\tau}\right)\left(-\mu\left(x_{1-\tau}\right)+\frac{1}{p\left(x_{1-\tau}\right)} \partial_{x_{1-\tau}}\left[\sigma^2\left(x_{1-\tau}\right) p\left(x_{1-\tau}\right)\right]\right)\right]+ \\
& \frac{1}{2} \partial_{x_{1-\tau}}^2\left[p\left(x_{1-\tau}\right) \sigma^2\left(x_{1-\tau}\right)\right]
\end{aligned}
$$

which finally gives us a stochastic differential equation analogous to the Fokker-Planck/forward Kolmogorov equation that we can solve backward in time:

$$
d X_\tau=\left(-\mu\left(x_{1-\tau}\right)+\frac{1}{p\left(x_{1-\tau}\right)} \partial_{x-\tau}\left[\sigma^2\left(x_{1-\tau}\right) p\left(x_{1-\tau}\right)\right]\right) d \tau+\sigma\left(x_{1-\tau}\right) d W_\tau \tag{7.9}
$$

Where $\widetilde{W_t}$ is the Wiener process that flows backward in time. By keeping the $\sigma^2(x_t)$ constant and independent of x_t and applying log-derivative trick, the drift simplifies to:

$$
\begin{aligned}
dX_\tau &= \left(-\mu(x_{1-\tau}) + \frac{\sigma^2}{p(x_{1-\tau})}\partial_{x-\tau}\left[p(x_{1-\tau})\right]\right)d\tau + \sigma(x_{1-\tau})\,dW_\tau, \\
dX_\tau &= \left(-\mu(x_{1-\tau}) + \sigma^2\frac{\partial_{x-\tau}\left[p(x_{1-\tau})\right]}{p(x_{1-\tau})}\right)d\tau + \sigma(x_{1-\tau})\,dW_\tau, \\
dX_\tau &= \left(-\mu(x_{1-\tau}) + \sigma^2\partial_{x-\tau}\log p(x_{1-\tau})\right)d\tau + \sigma(x_{1-\tau})\,d\widetilde{W}_\tau. \qquad (7.10)
\end{aligned}
$$

Where $\tau = 1 - t$

Notice that it's the very end at which the magic happens which allows us to formulate Score Based models.

8 Numerical Methods

8.1 Euler's method

The SDEs that we have seen until now account for the very small fraction of all possible stochastic equations for which an analytical solution is possible. For all others, one must solve these equations by simulating them using a computer (a process referred to as numerical simulation). To solve the stochastic equation:

$$d\boldsymbol{x} = f(x,t)dt + g(x,t)d\boldsymbol{W} \rightarrow (1)$$

in this way, we first choose an initial value for x . We can then approximate the differential equation by the equation:

$$\Delta\boldsymbol{x} = f(x,t)\Delta t + g(x,t)\Delta\boldsymbol{W} \rightarrow (2)$$

where $\Delta\boldsymbol{t}$ is a small fixed value, and $\Delta\boldsymbol{W}$ is a Gaussian random variable with zero mean and variance equal to Δt. For each time-step Δt, we calculate Δx by using a random number generator to pick a value for $\Delta\mathbf{W}$. We then add Δx to to x , and repeat the process, continuing to increment x by the new value of Δx for each time-step.

The result of the numerical simulation described above is an approximation to a single sample path of $\boldsymbol{x}$. To obtain an approximation to the probability density for x at some time T , we can repeat the simulation many times, each time performing the simulation up until time T, and each time using a different set of randomly chosen values for the ΔW s. This generates many different sample paths for x between time 0 and T , and many different samples of the value of x at time T . We can then make a histogram of the values of x, and this is an approximation to the probability density for x(T).We can also obtain an approximation to the mean and variance of x(T) simply by calculating the mean and variance of the samples we have generated.

The above method works just as well for a vector stochastic equation for a set of variables $\overrightarrow{\boldsymbol{x}} = (x_1, \ldots x_N)$, driven by a vector of independent noise increments $d\boldsymbol{W} = (dW_1, \ldots dW_M)$:

$$d\boldsymbol{x} = f(x,t)dt + G(x,t)d\boldsymbol{W} \rightarrow (3)$$

Where G is an Nx M matrix. As above the approximation to dx is given by replacing dt with Δt, and $\mathbf{dW}$ with $\Delta \boldsymbol{W} = (\Delta W_1, \ldots, \Delta W_M)$
This method for simulating stochastic differential equations is the stochastic equivalent of Euler's method for deterministic differential equations

Generating Gaussian Random Variables:

To realize the simulation described above, we need to generate Gaussian random variables with zero mean. The following method generates 2 independent zero mean Gaussian variables with variance $\sigma^2 = 1$. We first take two random variables x and y that are uniformly distributed on the interval $[0, 1]$ (All modern programming languages include inbuilt functions to generate such variables.) We then calculate:

$$\begin{cases} x' = 2x - 1 \\ y' = 2y - 1 \end{cases} \rightarrow (4)$$

These new random variables are now uniformly distributed on the interval $[-1, 1]$ rather than the interval $[0, 1]$ like they were before. We now calculate:

$$r = x'^2 + y'^2 \rightarrow (5 \tag{5}$$

If $r = 0$, or $r \geq 1$, then we return to the first step and calculate new random variables x and y . If $r \in (0, 1)$ then we calculate,

$$\begin{cases} g_1 = x'\sqrt{-2\ln(r)/r} \\ g_2 = y'\sqrt{-2\ln(r)/r} \end{cases} \rightarrow (6)$$

The variables g_1 and g_2 are Gaussian with zero mean and unit variance, and mutually independent. If instead we want g_1 and g_2 to have variance c , then we simply multiply them by $\sqrt{c}$.

Checking the accuracy of a solution

The accuracy of the sample paths, means, and variances that we calculate using Euler's method will depend upon the size of the time-step Δt, and the values of x , f and g throughout the evolution. The smaller the time-step with respect to these values, then the more accurate the simulation. The approximate sample path given by the simulation converges to a true sample path in the limit as the time-step tends to zero

How to check for the simulation:

A simple way to check the accuracy of a simulation that uses a particular time-step is to perform the simulation again, but this time with the time-step halved. If the results of the two simulations differ by only a little, then the first simulation can be assumed to be accurate to approximately the difference between them. This is because we expect the result of halving the time-step a second time to change the result much less than it did the first time.

The process of halving the time-step, Δt, of a simulation deserves a little more attention. Note that a given sample path is generated by a given realization of the noise, and this is the set of the ΔWs (chosen at random) for the simulation. Let us say that there are N time-steps in our simulation, and denote each of the N noise increments by ΔW_n, where $\mathrm{n} = 0, \ldots, \mathrm{N} - 1$. If we wish to halve the time-step, and generate an approximation to the same sample path, then we need to generate a set of 2 N Gaussian random numbers, $\Delta\widetilde{W}_m$, that agree with the previous set of N random numbers ΔW_n. What this means is that the sum of the first two stochastic increments, $\Delta\widetilde{W}_0$ and $\Delta\widetilde{W}_1$ must equal the first stochastic increment ΔW_0. This is because the total stochastic increment for the second simulation in the time-period Δt is $\Delta\widetilde{W}_0 + \Delta\widetilde{W}_1$ and this must agree with the stochastic increment for the first simulation in the same time-step, in order for the two simulations to have the same noise realization. This must also be true for the second pair, $\Delta\widetilde{W}_2$ and $\Delta\widetilde{W}_3$ etc.

So we require:

$$\Delta\widetilde{W}_{2n} + \Delta\widetilde{W}_{2n+1} = \Delta W_n, \mathrm{n} = 0, 1, \ldots \ \mathrm{N} - 1 \rightarrow (7)$$

An example of two approximations to the same sample path, one with half the time-step of the other, is shown in the figure below:

Figure 6.1. The solid line is a numerical approximation to a sample path of the Wiener process, with a time-step of $\Delta t = 1$. The dashed line is an approximation to the same sample path, but with half the time-step. For these approximations to correspond to the same sample path, the second must agree with the first at every multiple of Δt, denoted by the circles.

Fortunately it is very easy to generate a set of $\Delta\widetilde{W}_{2m}$ for which Eq. (7) is true. All one has to do is generate N random numbers r_n with mean zero and variance $\frac{\Delta t}{2}$, and then set:

$$\begin{cases} \Delta\widetilde{W}_{2n} = r_n \\ \Delta\widetilde{W}_{2n+1} = \Delta W_n - r_n \end{cases} \rightarrow \qquad (8)$$

The above procedure allows one to perform two simulations of the same sample path for an SDE with different time-steps. If the difference between the final values of x for the two simulations are too large, then one can halve the time-step again and perform another simulation. One stops when the process of halving the time-step changes the final value of x by an amount that is considered to be small enough for

the given application.

By repeatedly halving the time-step, one can also determine how rapidly the simulation converges to the true value of $x(\ T)$. The faster the convergence the better, and different numerical methods have different rates of convergence. The simple Euler method that we described above has the slowest rate of convergence.

The accuracy of a numerical method:

The accuracy of a numerical method is referred to as the order of the method, and for stochastic simulations, by convention, this order is quoted as one-half less than the power by which the error scales. The Euler method is thus said to be half-order. There is a perfectly sensible reason for this terminology. It comes from writing the approximate increment as a power series in Δt, which one imagines to agree with the true increment up to some particular power, and then disagree for higher powers. Because the ΔW part of the increment is proportional to $\sqrt{\Delta t}$ one must use a power series in $\sqrt{t}$. Thus if the error of the increment Δx is proportional to Δt as in Euler's method, then the increment is accurate to $\sqrt{\Delta t}$ (half-order in Δt), being the order of the first term in the power series. For a rigorous justification of these arguments, we refer the reader to the comprehensive text on numerical methods for SDEs by Kloeden and Platen [21].

In the next section we describe a simple numerical method that is accurate to first-order in t, and thus for which the error scales as $(\Delta t)^{\frac{3}{2}}$

Milstien's method:

The usefulness of Milstien's method comes from its simplicity, and the fact that it gives a significant improvement over Euler's method, since it is accurate to first-order in Δt. The Milstien method approximates the increment, dx , of the differential equation:

$$dx = f(x,t)dt + g(x,t)dW \rightarrow (9)$$

By,

$$\Delta x = f(x,t)\Delta t + g(x,t)\Delta W + \frac{g(x,t)}{2}\frac{\partial g(x,t)}{\partial x}\left[(\Delta W)^2 - \Delta t\right] \rightarrow \qquad (10)$$

Vector Equation with scalar noise:

For the vector differential equation with a single (scalar) noise source, dW, given by:

$$d\boldsymbol{x} = f(\boldsymbol{x},t)dt + g(\boldsymbol{x},t)dW \rightarrow (11)$$

Where $x = (x_1, \ldots x_N)^T$, $f = (f_1, \ldots f_N)^T$ and $g = (g_1, \ldots g_N)^T$. Milstein's Method is,

$$\Delta x_i = f_i \Delta t + g_i \Delta W + \frac{1}{2} \sum_{j=1}^{M} g_j \frac{\partial g_i}{\partial x_j} \left[(\Delta \text{ W})^2 - \Delta t \right] \rightarrow \tag{11}$$

Two special cases of this are as follows:

1. The Ornstein-Uhlenbeck process with scalar noise. The equation for this process is $\Delta x = Fxdt + gdW$, where F is a constant matrix, and g is a constant vector. As a result Milstien's method is just the same as Euler's method, being

$$\Delta x = Fx\Delta t + g\Delta W$$

This means that for additive noise, Euler's method is a first-order method.

2. The general linear stochastic equation with scalar noise. For the general linear stochastic equation with a single noise source, $\mathbf{dx} = F\mathbf{x}dt + G\mathbf{x}dW$, where F and G are constant matrices, Milstien's method becomes:

$$\Delta x = Fx\Delta t + Gx\Delta W + \frac{1}{2} G^2 x \left[(\Delta W)^2 - \Delta t \right]$$

Runge-Kutter-like methods:

A potential disadvantage of the Milstien method is that one must evaluate the first derivative of the function g(x, t) that multiplies the stochastic increment. For deterministic differential equations it is the Runge-Kutter family of methods that eliminate the need to evaluate such derivatives. Similar methods can be found for stochastic equations. Here we present a first-order method of this type that was obtained by Platen, building upon Milstien's method. We will refer to it as the Milstien-Platen method, and it is obtained from Milstien's method above by replacing the derivative of g with an approximation that is valid to first-order. For a stochastic equation containing only a single variable x , the first-order approximation to the term g that appears in Milstien's method is:

$$g(x,t) \frac{\partial}{\partial x} g(x,t) \approx \frac{1}{\sqrt{\Delta t}} [g(q,t) - g(x,t)] \rightarrow \tag{12}$$

With,

$$q = x + f\Delta t + g\sqrt{\Delta \text{t}} \rightarrow \tag{13}$$

Substituting this into Milstien's method for a single variable, Eq. (10), we obtain the MilstienPlaten method for a single variable:

$$\Delta x = f\Delta t + g\Delta W + \frac{1}{2\sqrt{\Delta t}} \left[g(q,t) - g(x,t) \left[(\Delta W)^2 - \Delta t \right] \right] \rightarrow \tag{13}$$

For a vector stochastic equation with scalar noise (Eq. (11)), the Milstien-Platen method is

$$\Delta x_i = f_i \Delta t + g_i \Delta W + \frac{1}{2\sqrt{\Delta t}} \left[g(\boldsymbol{q}, t) - g(\boldsymbol{x}, t) \left[(\Delta W)^2 - \Delta t \right] \right] \rightarrow (13)$$

where the ith element of the vector $\mathbf{q}$ is:

$$q_i = x_i + f_i \Delta t + g_i \sqrt{\Delta t} \rightarrow \ (14)$$

9 Limitations and Open Challenges

Discussion of current limitations and potential challenges for future research in these domains.

A Appendices

A.1 Code Resources

Links to relevant GitHub repositories, tools, or resources.

A.2 Suplementary Material

Supplemental proofs and derivations supporting the discussed methodologies.

A.2.1 Tweedie's Formula

A.2.2 Total Variation Regularization

A.3 Additional Figures

Figures and diagrams for extended illustration of key concepts.

Acknowledgments

This work has been supported and guided by Dr. Hassan Mohy-ud-Din and Dr. Muhammad Tahir. Their mentorship and expertise have been instrumental in developing this research.

References